

Facing reality at NASA

The US space agency is facing a crisis that neither it nor its political stakeholders seem willing to tackle. Something big has to give, and it shouldn't be science.

Some will say that scientists overreacted to last week's threatened termination of NASA grants for outer-planets research (see page 447). It was much ado about nothing — the money will be awarded after all. But it shows how jumpy space scientists are these days at any hint that their budget might be raided.

They're right to worry. NASA is heading for deep financial trouble, and no one in Washington appears willing to say so.

NASA scored a victory of sorts in the autumn when the White House and Texas Republican Tom DeLay, the most powerful figure in the House of Representatives, made sure the agency got nearly its full budget request of \$16 billion. Congress also removed the barriers that prevented NASA managers from using space-science funds to solve other problems within the agency.

But these moves can only temporarily hide the truth — that NASA simply can't afford a vigorous space-science programme, a space station, the space shuttle and a new Moon–Mars astronaut programme, all on \$16 billion a year. The shuttle costs \$4 billion a year when things are going smoothly. NASA plans to fly 28 shuttles to the space station between now and 2010 — more than the average of 4.7 launches a year, even though the ageing shuttle is down from four vehicles to three and launch directors will be more cautious than ever. Expect delays and cost overruns: the cost of fixing a few systems after the Columbia accident is \$1.5 billion and counting. The White House promised that its new "vision" would be paid for by phasing out the shuttle and space station, but that already seems like a pipe-dream.

If NASA is committed to finishing the space station, and the Moon–Mars plan remains a White House priority, that leaves only

science to cut. But agency managers won't admit this. The battle over servicing the Hubble telescope is a case in point. Outgoing administrator Sean O'Keefe claimed that his decision last year to ban astronaut repair missions to Hubble was based solely on safety. But a National Academy of Sciences panel and the head of the Columbia accident investigation both say a Hubble mission is not significantly more dangerous than a trip to the space station.

No wonder scientists suspect it's really about money. NASA says that any Hubble servicing option — using astronauts or robots — will cost about \$2 billion. What do they get for this? A November report by the Government Accountability Office revealed that nearly half the estimated price of an astronaut servicing mission is for extending the shuttle's life by three months beyond the planned phase-out in 2010. In other words, NASA would charge its space-science office for the repair flight, something it never used to do, and had not planned to do before the Moon–Mars plan was announced.

Congress will take up the subject of Hubble servicing in a hearing this week. Politicians in Maryland, the home state of the Space Telescope Science Institute, have already promised to come to the telescope's rescue, and some astronomers hope Congress will simply restore the funds for Hubble servicing if the White House cuts them from the NASA budget as expected.

But that would be no substitute for honesty and realism. If NASA maintains its present course, the shuttle and space station will be a drain on its finances for years, leaving little money for the Moon and Mars. And it will be an even greater tragedy if science, NASA's most forward-looking enterprise, suffers to make up the shortfall. ■

Turf battles versus German excellence

Disputes between Germany's states and its federal government need not cripple the country's research priorities.

When Chancellor Gerhard Schröder last month opened Germany's celebrations of Albert Einstein's great achievements in 1905, he expounded at great length the urgency of boosting creativity, innovation and enthusiasm for science. But he failed to mention the fact that an ambitious attempt to boost support for curiosity-driven research in Einstein's country of origin was about to crumble, thanks to political manoeuvring.

The programmes designed last year to create world-class centres in German universities and research institutes may fall victim to the chronic battles for power between federal and regional governments (see page 448). Both sides had agreed to support the programmes, and the money — €390 million (US\$510 million) per year from 2006 to 2010 — had been set aside. Shamefully, the bickering has led to the programmes being put on ice.

The irony is that the impasse is a result of an attempt to undo the Gordian knot of federalism in Germany by loosening the jealously guarded powers of the regions. The attempt has failed dismally. Forlornly, one can only hope that the politicians will learn that science and other key areas are the ultimate victims of their power games.

At the moment, science is being used as a hostage. With Schröder's

praise of science still ringing in their ears, this will seem a cynical tactic to Germany's tens of thousands of highly motivated researchers.

But there may be a productive way forward. Some suspect that the *Länder* (states) governed by the Christian Democrats begrudge the Social Democrat federal government any political success: they may fear that the extra money for science will begin to bear fruit just before the next federal elections in 2006. But rather than block such success, they should try to get a share of it.

The fairest way would be for all to agree to entrust the DFG, Germany's main science funding agency, with the task of distributing in a transparent and competitive way the extra money in the framework of a new programme of research excellence. There is no lack of ideas about how to spend the extra money efficiently. For example, there are many groups in Germany for whom relatively modest equipment grants would enable competition with top centres overseas.

Investment in a few dozen interdisciplinary training programmes, such as the high-profile graduate school for cellular neuroscience at the University of Leipzig, would improve training and career opportunities for many young researchers and greatly increase Germany's attractiveness for the brightest minds from abroad. ■





Search terms
Publishers bring
Google to book over
digitized texts
p446



Lost and found
Los Alamos Lab
reveals missing disks
never existed
p447



Down to Earth
Geophysicists aim
for groundbreaking
initiative in Africa
p449



Fruitful delivery
Flies finally pass
the post in mail-
dominated world
p450

DNA is burning issue as Japan and Korea clash over kidnaps

David Cyranoski, Tokyo

A bitter dispute has erupted between Japan and North Korea over DNA tests used to establish whether cremated remains belong to a Japanese citizen abducted in 1977.

The argument is the latest twist in an episode that has soured relations between the two countries for years. During the 1970s and 1980s, North Korea was believed to have abducted at least 13, and perhaps as many as 100, Japanese citizens to work in its espionage programme. Now the two nations are falling out over the feasibility of correctly identifying DNA from the ashes of one of those abducted.

In the autumn of 2002, North Korea ended years of denials and admitted that members of its armed forces abducted 13 Japanese citizens from Japan and Europe. The North Korean leader, Kim Jong-il, claimed that the abductions were carried out by the military without government permission. Japan says it has evidence of two more abductees, and believes that there were in fact many more.

Pressure on North Korea to release survivors has since seen five abductees return to Japan accompanied by their families. Information on eight others, which North Korea says are dead, has been slow to arrive, leading to speculation that some of them are still alive.

On 15 November last year, Japanese officials returned from talks in Pyongyang carrying what North Korea claimed to be the cremated remains of Megumi Yokota, who was abducted in 1977 at the age of 13. According to North Korean reports, Yokota married a North Korean, but later killed herself after entering a mental hospital.

At Teikyo University in Tokyo, tests on five samples of the ashes found DNA from two sources — but neither of them matched DNA from Yokota's umbilical cord, which had been kept by her parents, as is common in Japan. In December, these results were passed to North Korea, but on 26 January the Korean government issued a statement that branded them a "fabrication".

According to Hatsuhsa Takashima, a spokesman for Japan's foreign ministry, North Korea called into question the methods used in the tests and claimed that the remains,



Missing: Megumi Yokota some years after her abduction, in a photo recently released by North Korea.

which had been heated to 1,200 °C, could not contain any surviving DNA. The North Korean statement also asked why researchers at Teikyo University were able to extract DNA when the National Research Institute of Police Science in Tokyo, which also had five samples to work with, had been unable to do so.

Sample survival

Teikyo University's Tomio Yoshii, one of Japan's leading forensics experts, says there are several reasons why he managed to extract DNA from all five of his samples. These include the fact that he used a highly sensitive process called the nested polymerase chain reaction (PCR), which amplifies DNA twice instead of once as in conventional PCR, and the possibility that his original samples were of better quality than those at the other lab. "Everyone has their own method" of handling DNA samples, he notes. "There is no standardization."

Little forensic work has been done on cremated specimens in Japan, and most experts, including Yoshii, thought it unlikely that DNA would have survived cremation at 1,200 °C. "I was totally surprised," says Yoshii. But DNA could survive if exposed to such heat for only a short time. "You can't tell

anything from temperature alone," says Hirofumi Fukushima, a forensics expert at Shinshu University.

Nonetheless Yoshii, who has no previous experience with cremated specimens, admits his tests are not conclusive and that it is possible the samples were contaminated. "The bones are like stiff sponges that can absorb anything. If sweat or oils of someone that was handling them soaked in, it would be impossible to get them out no matter how well they were prepared," he says.

Takashima says that North Korea has sent remains that it said belonged to an abductee in the past, only to admit later that they were from another source.

The Japanese government responded to the current incident on 26 January by calling North Korea's handling of the situation "deplorable". It threatened "stringent actions" that, according to Takashima, may include the cancellation of 125,000 tonnes of food aid and other trade sanctions.

Japanese officials also say that they want to retest the DNA in question. But Yoshii says his five samples — the largest of which weighed only 1.5 grams — were used up in his tests. And that, observers say, leaves little prospect of the disagreement being resolved. ■

NATL ASSOC. FOR THE RESCUE OF JAPANESE KIDNAPPED BY NORTH KOREA/AP

Publishers irritated by Google's digital library

Declan Butler, Paris

A spat is brewing between academic publishers and Google over the Internet-search company's plans to digitize and index library collections at major research universities.

Late last year, Google, based in Mountain View, California, announced a decade-long project to scan millions of volumes at the universities of Harvard, Stanford, Michigan and Oxford, as well as the New York Public Library. The resulting archive would allow computer users worldwide to search the texts online. But some publishers complain that they weren't consulted by Google, and that scanning library collections could be illegal.

Under the scheme, people searching with Google would find library volumes relevant to their query at the top of their search results. Clicking on a title would allow them to browse images of the full text of works in the public domain. Only brief excerpts and bibliographic data would be shown for material under copyright. Participating libraries would also be given a digital copy of their collection.

Google describes the initiative as an extension of Google Print (www.print.google.com), which is based on agreements with publishers and allows the full text of books to be searched. Google Print's results provide a brief excerpt of the text, together with a link to publishers or booksellers that sell the book and to libraries that hold it.

But Google has not yet struck any legal



Now booking: Google wants to digitize research libraries.

agreements with publishers, either individually or collectively, for the research-library initiative, says Sally Morris, chief executive of the Association of Learned and Professional Society Publishers, the international trade body for not-for-profit publishers. Few publishers would want to opt out of the library scheme, Morris says — but they need to be asked to provide the appropriate permission.

Copyright material generally carries some variation of a warning banning the reproduction, storage or distribution of copies of the work without the publisher's permission. Scanning a book constitutes making a copy and so is only allowed with permission, say lawyers from several publishers. They also argue that an exception under US law that allows libraries to copy texts for preservation purposes would not

apply in this case. Nor would making copies for 'fair use', given that Google is a commercial company.

A spokesman for Google says that it will "respect the rights of copyright holders", and that it "prefers to work directly with publishers to bring copyrighted books online". Google "has been working closely with publishers to help them connect with more readers online", he adds.

Part of the uncertainty stems from the fact that there seems to have been little discussion so far between Google and publishers, says Terry Hulbert, head of electronic development and strategy at the UK Institute of Physics. "Someone clearly needs to have a chat with the 800-pound gorilla sat in the corner," he observes. "There is no question that Google should have spoken to the learned societies and publishers beforehand. Systematic digitization of copyright content is absolutely something they cannot do without seeking approval of the rights holders."

Peter Kosewski, director of publications and communications at Harvard University Library, says the library believes that the way Google intends to handle copyright works is consistent with the law. Harvard is carrying out a pilot with Google on 40,000 titles before making a decision on digitizing its entire 15-million-volume collection. "We have a number of questions that will be answered by the pilot project, and that includes copyright issues," he says. "We think it is a great programme Google has put together." ■

University dispute puts Berlin science meetings in crisis

Alison Abbott, Munich

The future of Berlin's prestigious Dahlem Conferences is hanging in the balance this week, as members of their scientific advisory board contemplate how to end a stalemate with the Free University, which administers the meetings.

The 12-strong board of international scientists believes that the university is trying to muscle in on the reputation of the conferences to promote itself within the highly competitive German university system.

According to its chair, Randolph Menzel, a neurobiologist at the Free University, some board members now want to resign. But he says he still hopes that the board can resolve its differences with the university.

Dieter Lenzen, president of the Free University, told *Nature* that the university remains committed to the conferences, and does not want to change the way they are run.

The Dahlem Conferences were established in 1974 in West Berlin as part of a general plan to make the city — isolated in the former East Germany — more attractive to key international players in all areas of society.

For each meeting, a 'hot' scientific problem is selected by the advisory board, along with 40 or so participants, who break into groups to discuss four key questions related to the topic. Ninety-five meetings have been held so far, each followed by a publication.

After German reunification in 1990, Berlin lost its generous federal subsidies and the Free University had to find money for the conferences from its own budget. At the same time that budget was cut drastically — more than two-thirds of tenured academic positions have since disappeared. Last year, an International Advisory Committee (IAC) of investors and scientists was created to help raise additional funds.

But tensions soon arose between the IAC and the scientific advisory board. Norbert Baer, a member of the science board and a conservation expert at New York University, says that there are numerous examples of the Free University's administration interfering in the board's work. "Even the creation of the IAC without consulting us was an infringement of our role," he says.

A crisis point came when the Free University sacked the Dahlem Conferences series editor Julia Lupp last November, without consulting the board. Negotiations to try to get her reinstated failed last week.

"I think the Dahlem conferences have been so seriously compromised that they should be removed from the caretaking of the Free University and run with other scientific organizations in Berlin," Baer says.

Members of the scientific board meet on 4 February to consider their next move. ■

Lab relations sour as 'missing disk' charges are proved false

Geoff Brumfiel, Washington

Two classified computer disks that allegedly vanished last summer at the Los Alamos nuclear-weapons laboratory in New Mexico never existed, according to an investigation by the government agency that oversees the lab.

The security lapse, together with an unrelated accident, led to a three-month shutdown of the laboratory last summer, with director Peter Nanos accusing scientists of operating in a "cowboy culture" (see *Nature* 430, 387; 2004). The conclusion that the disks never existed has infuriated many of the lab's researchers.

"The talk in the halls is mutinous," says Doug Roberts, a computer scientist at the laboratory. "I've been at the lab for 20 years and morale has never been this bad before."

The Los Alamos National Laboratory has been battered in recent years by a wave of scandals. In 1999, it was the subject of national scrutiny when Wen Ho Lee, a Taiwanese-born scientist, was accused of smuggling nuclear secrets to China (see *Nature* 398, 96; 1999) and subsequently acquitted. In 2000, two computer hard drives containing classified data disappeared from a secure area inside the laboratory, only to reappear later behind a photocopier (see *Nature* 405, 725; 2000). And in 2003, the laboratory's director and deputy director resigned following accusations that they had improperly fired two whistleblowers who had alleged widespread theft at the lab (see *Nature* 421, 99; 2003).

The latest trouble for the laboratory began early last July, when an inventory of classified data in its weapons-physics directorate revealed that four disk drives were missing. Almost immediately, two of the drives were found to have been improperly moved to a different building, but another two could not be located. In response, Nanos shut down large parts of the laboratory and publicly chided the scientists working there for failing to follow security procedures. "This willful flouting of the rules must stop, and I don't care how many people I have to fire to make it stop," he wrote in the 2 August issue of the laboratory's newsletter.

But now it seems that the missing drives were in fact an artefact of flawed inventory procedures. According to the report by the National Nuclear Security Administration



Scold news: lab chief Peter Nanos's reproaches may spark mutiny.

(NNSA), which was released on 28 January, 12 barcodes used to catalogue classified disk drives were issued to a group that needed only 10. The extra barcodes were nevertheless included in a master list, and so when auditors conducted an inventory last July, they concluded that two disks were missing. "The allegedly missing disks never existed and no compromise of classified material has occurred," the report explains.

Many scientists at the laboratory say that the incident, together with Nanos's public rebuke, has profoundly damaged the relationship between Los Alamos researchers and the lab's management.

"Trust in upper management has been completely lost," says Brad Holian, who has worked as a theoretical physicist at the laboratory for 32 years. Holian says that the three-month shutdown was the breaking point for many already frustrated scientists. "We were told in the theoretical division that we couldn't write down calculations on the blackboard," he says. Many of his colleagues are leaving the lab, and Holian himself says that he plans to retire this March — years earlier than he had originally planned. "I think there are a lot of people in my situation," he says.

In a statement accompanying the report, NNSA administrator Linton Brooks said that the University of California, which oversees the laboratory, would be fined around \$5.1 million for what he describes as "major weaknesses in controlling classified material revealed by this incident". Chris Harrington, a spokesman for the university, says that some of the money for the fine will come from the laboratory's discretionary research budget.

Cancellation e-mail shakes recipients of outer-planet grants

Tony Reichhardt, Washington

In the end, it was put down to a misunderstanding. But dozens of planetary scientists were enraged last week to be told — erroneously, as it turned out — that their NASA grants had been terminated.

Curt Niebur, who administers the US space agency's research programme for the outer planets, wrote to grantees on 24 January saying: "I regret to inform you that the [fiscal year 2005] funds to support this program have been redirected by the order of the NASA Administrator to meet other agency needs."

The money, which totals \$4.8 million for 55 grants, funds the analysis of data from missions such as Galileo and Cassini, as well as the basic research that lays the foundations for future missions.

Niebur's letter prompted a flurry of e-mails and an immediate reaction from the Division for Planetary Sciences of the American Astronomical Society. Grantees were asked to contact Congress, and to write testimonials about the negative effects of suddenly losing their funding. One postdoc wrote: "I just moved across the country to work. Now I have no job."

Almost immediately, NASA science managers began back-peddalling, claiming that the problem had been blown out of proportion. On 27 January, Andrew Dantzler, acting director of the agency's Solar System division, wrote to the grantees: "Your Outer Planets Research funding has not been cut." Instead, he wrote, it was being moved to another account "for purely administrative reasons".

It was all due to miscommunication within his office, Dantzler says — and it won't happen again. The office has no plans to cut grant money, he adds, calling his research funding "quite healthy".

But Dantzler acknowledges that "the timing couldn't have been worse", coming just as scientists fear the imminent cancellation of NASA's Jupiter Icy Moons Orbiter mission (see *Nature* 433, 342; 2005). The flap demonstrates just how edgy space researchers have become, with every day bringing fresh rumours of cuts to NASA science programmes, such as servicing of the Hubble telescope.

In their initial anger, some scientists were clearly prepared to think the worst of NASA. One anonymous e-mail said: "This administration's attitudes toward basic science and education are extremely disappointing."

Political deadlock delays promised German research cash

Quirin Schiermeier, Munich

Germany's top science manager is urging the federal and state governments to go ahead with two funding programmes that have been grounded by a political impasse.

"Science in Germany has become a hostage of a political power game," Ernst-Ludwig Winnacker, the president of the DFG, Germany's main funding agency for basic research, told *Nature*. "We've got to get out of this unprecedented deadlock."

The programmes consist of a €2-billion (US\$2.6-billion) scheme to create several elite universities and a separate 'pact for research' that would guarantee 3% annual budget increases until 2010 for the Max Planck Society and other large non-university research bodies.

The projects have been frozen since December, when the planned reform of the relationship between Germany's federal government in Berlin and its 16 *Länder* (states) fell apart. The plan failed largely because the Social Democrat-led government and the Christian Democrats, who control most of the *Länder*, disagreed about who should have power over science and education. Science experts from the federal government and the *Länder* are now discussing the possibility of going ahead with the programmes anyway.

"Science has without doubt climbed up the political agenda in Germany," says Winnacker. "Unfortunately, this appears to have caused more harm than good."

The impasse is hurting Germany's universities, which carry out four-fifths of the country's publicly funded research. In January last year, Edelgard Bulmahn, the science minister, promised to award a small group of elite universities an extra €390 million each year for five years from 2006 (see *Nature* 427, 477; 2004).

Now candidate institutions are growing impatient as they await confirmation that the money will be released. Patrick Cramer, for example, a structural biologist and managing director of the University of Munich's Gene Centre, says he needs the money to attract top researchers. "We cannot yet compete with the best groups in the United States in terms of equipment," he says.

Although an end to the impasse is currently not in sight, Winnacker remains optimistic that sooner or later the promised money will flow. Disagreement over who should distribute the extra money, and how, should not threaten the programmes as a whole, he says. ■



Graham Watkins' appointment could ease tensions for researchers at the Charles Darwin Foundation.

Hopes rise as head named for troubled conservation centre

Henry Nicholls

Ecologist Graham Watkins has been named executive director of the Charles Darwin Foundation (CDF), the scientific authority in the Galapagos archipelago. Observers hope he will bring some much-needed stability to the islands' conservation effort.

The past decade has been difficult for scientists in the Galapagos, as pressures on its natural resources have become intense. Oil spills, hunting and fishing have all taken their toll — and researchers say that fishermen, frustrated by quotas, have made death threats against them, seized research buildings and forced some facilities to close (see *Nature* 408, 761; 2000). Despite this, the cluster of islands off the coast of Ecuador remain of intense interest to researchers studying evolution, as they have done since Charles Darwin's visit on the *Beagle* 170 years ago.

The situation has not been helped by problems within the foundation in recent years. "There was a general loss of confidence in its leadership," says Nigel Sitwell, chairman of the Galapagos Conservation Trust, the UK-based charity that raises money for conservation in the islands. The board of directors was completely remodelled in January 2004, he says, and the foundation has been held together by an acting director since then. "Strong leadership and fresh ideas should help to ease the current tensions," says Sitwell.

Watkins was due to take up office in the Charles Darwin Research Station on

1 February, returning to the islands where he was a tourist guide in the 1980s. "The islands have left an indelible mark in my mind," he says. Watkins comes from a two-year post as director-general of the Iwokrama International Centre for Rain Forest Conservation and Development in central Guyana.

Part of Watkins' remit will be to secure more money for Galapagos conservation. But the first major challenge, he says, will be to strengthen the dialogue between the CDF and the Galapagos National Park Service, fishermen, naturalist guides and the tourism industry. In particular, the partnership between the CDF and its sister body, the national park service, is crucial, he says.

But the national park service has troubles of its own. Since 2002, leadership of the park service, appointed by the Ecuadorean government, has changed hands about eight times. Last September, wardens went on strike to protest against the latest incumbent, Fausto Cepeda, saying that he had ties with the fishing industry. The park service is still awaiting the outcome of an independent selection process to find a new director.

Sitwell says that Watkins' appointment to run the CDF is a step in the right direction. "The Galapagos is one of the most important protected areas in the world," says Watkins. "Its conservation and effective management are critical not only to its own future but as a model for other protected areas." ■

African network set to boost Earth sciences

Rex Dalton, San Diego

Geophysicists are joining forces behind a programme called AfricaArray, which they hope will spread the training and practice of geophysics throughout Africa.

Many African economies are heavily dependent on oil and mineral extraction — ensuring that skills in geophysics are in strong demand. Yet young, trained geophysicists remain in short supply across the continent, say companies involved in energy and mining.

That is one reason why Andrew Nyblade, a geophysicist at Pennsylvania State University, University Park, decided to initiate AfricaArray. The project plans to channel funds from the public and private sectors into research and education in geophysics.

Nyblade, who was born in Tanzania, says the idea is to combine first-rate research, backed by top agencies such as the US National Science Foundation, with training for local specialists who will stay and work in Africa. “The brain drain is a big issue,” he says.

The programme — established as a partnership between Pennsylvania State University, the South African Council for Geoscience in Pretoria and the University of the Witwatersrand in Johannesburg — has so far attracted more than US\$600,000 of the \$2.7 million that it hopes to raise.

AfricaArray researchers will focus on



Array of hope: Andrew Nyblade (centre) wants more African geophysicists.

important seismological and volcanic activity in Africa. For example, a network of seismometers will be deployed to examine the structure of the superplume of magma beneath the continent.

A boom in the price of rare mineral commodities, including gold, is currently fuelling demand for trained Earth scientists in Africa. “The Earth sciences play major roles in the mapping of important minerals,” says Gerhard Graham, head of scientific services at the Council for Geoscience. “But lack of basic geological knowledge limits development in large parts of our mineral-rich continent.”

Graham says he hopes that AfricaArray “can become a driving force, in particular for countries in eastern and southern Africa to

participate in the development of Earth sciences within their own countries.”

Paul Dirks, head of geosciences at Witwatersrand, says that more global cooperation is needed. “Geophysical training programmes in Africa are not widespread or strong,” says Dirks. “We have to go international to make them sustainable.”

Dirks and Nyblade have just completed a round of US visits to multinational corporations, including ChevronTexaco, Exxon-Mobil and Schlumberger, to seek more support for AfricaArray.

Nyblade is already using a \$165,000 grant from the South African government to rework an existing net of 11 seismometers. With 21 additions, this will be used to determine the depth of the superplume and study its role in heat convection from the inner Earth. “To me, this is the big research prize,” says Nyblade.

Mark van der Meijde, a geophysicist at the International Institute for Geo-Information Science and Earth Observation in Enschede, the Netherlands, says that AfricaArray will “open up unprecedented possibilities for starting new projects”.

Geophysicists will hold a workshop in Palmanova, Italy, on 26–27 February, to refine their plans for the programme.

Additional reporting by Quirin Schiermeier in Munich.

► <http://africaarray.psu.edu>

Reformation of bird-brain terminology takes off

Jessica Ebert, Washington

Neuroscientists are revamping a naming system for birds’ brains that has been in use for more than a century.

The old terminology hinders communication between bird neuroscientists and their mammalian counterparts, avian specialists say, because it does not reflect modern understanding.

The problem dates back to Ludwig Edinger, a German neuroscientist working in the nineteenth century. Edinger thought the cerebrum of the bird brain was primitive and consisted of nothing more than basal ganglia that control instinctive behaviour. In contrast, he thought, the mammalian brain consisted of layers that create a ‘neocortex’ and control learning.

Neuroscientists have known for decades that such a distinction is artificial. Signalling molecules and neurotransmitters operate similarly in the brains of birds and

mammals. And researchers agree that birds can learn: crows can pass on tool-making skills, for example.

But the different terminology meant “the value of avian research was not coming across to mammalian researchers”, says Anton Reiner, a neuroscientist at the University of Tennessee Health Science Center at Memphis.

So a new system has been designed by the Avian Brain Nomenclature Consortium, a group of 29 specialists in bird, fish, reptile and mammalian brains, who first met at Duke University in Durham, North Carolina, in July 2002.

The meeting led to a naming system that illuminates the parallels between bird and mammalian brains. A technical paper describing the terminology was published last May in the *Journal of Comparative Neurology* (A. Reiner *et al.* 473, 377–414; 2004). It was quickly adopted by “the entire

avian community”, says Erich Jarvis, a neuroscientist at Duke University who hosted the meeting. This month’s *Nature Reviews Neuroscience* (E. Jarvis *et al.* 6, 151–159; 2005) carries a general introduction to the system for the wider community.

Stephanie White, a neuroscientist at the University of California, Los Angeles, says the terminology helps her to compare vocal learning behaviour in birds and mammals. “I don’t even refer to the old nomenclature anymore,” she says.

But some argue that the terminology does not go far enough. A nomenclature “that is in greater harmony with mammalian terms and concepts”, remains possible, says George Paxinos, a neuroscientist at the Prince of Wales Medical Research Institute in Randwick, Australia. Until then, however, he says that he is grateful for the work the consortium has done to “get us out of the mess we were in”.

Indian Ocean nations seek regrowth to recover from tsunami

New Delhi With discussions under way for an early warning system for tsunamis in the Indian Ocean, the countries affected by the 26 December disaster are pressing ahead with projects to heal and protect the landscape damaged by the giant waves.

The International Rice Research Institute in the Philippines, for example, has shipped six varieties of salt-resistant rice to Malaysia and Sri Lanka, the countries whose rice fields were hardest hit. It also plans to send two researchers to the area in February to investigate how the land can best be reclaimed.

The Indian state of Tamil Nadu, meanwhile, has announced plans to build a 300-metre-deep plantation along its entire 1,000-kilometre coastline to act as a buffer against future storms and tsunamis. It plans to grow coconut and casuarina, which has been used before to stop coastal erosion. Officials say that the planting will be completed in three months, and that they expect other coastal states to follow suit.

Riot police called in to deal with rise in bird flu cases

London As cases of avian flu continue to spring up among birds in Vietnam, local officials have deployed riot police at checkpoints around Ho Chi Minh City. The armed officers are trying to stop infected or uncertified birds from being brought into the city, officials say.

As of 28 January, the World Health Organization reports that nine people have died from the H5N1 bird flu virus in Vietnam since mid-December. More recent news reports put the figure at 12.

One of the most recent victims is a 13-year-old girl, whose mother also died



Flying squad? Constant checks in the markets aim to keep avian flu out of Ho Chi Minh City.

Special UN delivery rescues illicit fly posters

Munich After years of sending live fruitflies through the international mail illegally, researchers will soon be able to post them while sticking to the letter of the law.

Research institutions in some 50 countries use fruitflies for various biomedical and genetic studies, and labs have long shipped them using regular post. But technically this has been illegal, according to the conventions of the United Postal Union (UPU), the Bern-based United Nations agency that oversees international mailing. The only animals allowed to be shipped live were leeches, bees and silkworms (see *Nature* **429**, 791; 2004).

After protests from researchers, the UPU decided last autumn to add to that list the family Drosophilidae, which includes fruitflies. The rule change was due to come into effect on 1 January 2006, but has now been brought forward to 1 May 2005.



The question of whether genetically modified fruitflies, or other research organisms such as the nematode *Caenorhabditis elegans*, can also be posted is still to be settled.

from the H5N1 virus. Vietnamese scientists are trying to determine whether this was a case of human-to-human transmission. The authorities are concerned that the virus could evolve to spread more easily between people, causing a pandemic.

Health experts suspect that there may also be human cases in neighbouring Cambodia and Laos, where there is minimal disease surveillance. One Cambodian woman is suspected of having died from the virus.

Danish ecologist is 'presumed innocent'

Paris A French committee investigating allegations of 'scientific dishonesty' against behavioural ecologist Anders Pape Møller has reported that it has insufficient evidence to prove either his innocence or his guilt. The committee of four independent scientists was established last year by the CNRS, the French national research agency.

In October 2003, the Danish Committees on Scientific Dishonesty ruled that a 1998 research paper retracted by Møller from the journal *Oikos* contained results that "must, at least in part, [have been] fabricated" (see *Nature* **427**, 381; 2004). Møller, a Dane now working at the CNRS Laboratory of Evolutionary Parasitology in Paris, had been unable to provide original data for that inquiry. He denied the allegation.

The ruling proved controversial and the CNRS decided to carry out its own investigation. The CNRS committee says that Møller did not follow good laboratory practice, particularly in the archiving of biological materials and data. But the committee also says it did not obtain enough material to make a clear judgement, so it concludes that Møller should be presumed innocent.

Africa takes centre stage for economic aid and ethics

Paris Africa's health and economic plight was in the spotlight last week in Davos, Switzerland, as world leaders, captains of industry and movie stars met for the annual World Economic Forum. They agreed to make helping Africa a top priority, and many pledged money to the cause.

Elsewhere, moves were being made to support African science. In Paris last week it was announced that 15 African countries are teaming up with the World Health Organization and medical research agencies in France, Britain and Germany to ensure that their science is done with proper ethical oversight. The project — Networking for Ethics on Biomedical Research in Africa — is expected to attract more research and clinical trials to the countries that need them most.

Gem of a writer set to head brain trust

Washington William Safire, a conservative US commentator well known for his political column in *The New York Times*, is to focus his energies elsewhere.

On 31 January, 75-year-old Safire took up the position of chief executive of the Washington-based Dana Foundation, which funds research and encourages interdisciplinary dialogue in neuroscience, immunology and the arts. He began working with the foundation a decade ago.

President George W. Bush and most conservative pundits have fought to restrict federal funding for stem-cell research, but Safire is a proponent of such studies. "I don't envisage the Dana Foundation ever lobbying for stem-cell research," says Safire, "but we will provide forums for discussions on the neuroethics of the issues."



To love, honour and assay: chemist Shazia Anjum had a clause put in her marriage contract allowing her to continue working.

Women at work

Pakistan's traditional ways have blocked many women's careers in science. But, as Ehsan Masood discovers, women are now fighting for their rights, both in life and in research.

“When I was doing my PhD, my husband and children would come to the department if I wanted to stay after 9 p.m.,” recalls Zahida Maqsood, a professor of chemistry at the University of Karachi. This wasn’t just to keep her company; nor to hassle her to come home and cook. In most Pakistani families, women are simply not allowed to work late alone, or to socialize after work with male colleagues. Even today, some 25 years since Maqsood’s graduation, things are much the same. “If my students want to stay late, their families come and wait for them outside on the lawn,” says Shakeel Farooqi, assistant professor of genetics at Karachi.

But the winds of change are beginning to stir. Secular women’s liberation movements have had very little impact in Pakistan. But the country, like many predominantly Muslim states, is witnessing the birth of Islamic

feminism, in which women are demanding rights that they say Islam accords to them. It is being fuelled in part by more women going to schools and universities, and also by a generation of women evangelists such as Farhat Hashmi, founder of the worldwide Al-Huda International Welfare Foundation, which teaches women about Islam. Hashmi’s lectures draw crowds of up to 10,000 women.

Employment rights

The effects are being felt throughout the country — including in its labs. Chemist Shazia Anjum, for example, is using her Muslim marriage contract to ensure that her science career moves forward. When she married she insisted upon a ‘right to work’ clause that would prevent her husband or his family from stopping her getting a job. Such legally binding, bespoke clauses are allowed in Muslim marriages, and can be inserted by

the bride or groom; a woman may ask to be paid for her housework, for example. But a combination of low levels of female literacy and established social conventions mean that few of Pakistan’s women are aware of these rights. Even fewer would think of using them. For Anjum, it has made possible her assistant professorship in chemistry at the International Centre for Chemical Sciences at the University of Karachi.

Some, including a former head of Pakistan’s Medical Research Council, claim to have had even fewer problems advancing their professional lives. Tasleem Akhtar comes from the North-West Frontier province, which is governed by an alliance of Islamic parties and is better known in the West for its role in nurturing the Taliban. Yet she still made it to the top job in medical research without, as she puts it, “needing to play golf with the minister for health”. She claims never to have felt the cold pressure of a glass ceiling; nor did she find opposition from within her own family. “I was allowed to study in the United Kingdom, no problem,” she says. “Instead of being forced to marry, daughters from poor families are often encouraged to get an education and earn a living for the rest of the family. It is in lower- and upper-middle-class families where you find the pressure to marry instead.”

Although Anjum and Akhtar’s experiences are a sign of changing times, many cultural barriers remain. Women rarely live alone in Pakistan, unless they are widowed or

divorced. Most live instead in extended families, in which parents, husbands and brothers often have the final say in decisions affecting their lives. Women cannot apply for a university place, a public-sector job or even a passport without completing a section on an application form that asks them for details of their father or husband. Most are forbidden by families to travel long distances without a male escort. Few drive cars beyond the limits of a handful of large cities and none ride bicycles or motorbikes on public roads. Women cannot marry without the written consent of a male, usually their father.

Liberal attitude

This kind of life may look painfully restrictive to Western eyes, but women's rights is now a cause with top-level backing from no less than President Pervez Musharraf and the chairman of Pakistan's Higher Education Commission, Atta-ur-Rahman, who was Anjum's PhD supervisor. Some of this support comes in the form of minor initiatives, such as encouraging higher-education institutions to provide women-only transport to ferry them safely home late at night. Other actions are much more significant. Musharraf, for example, has piloted a new law that formally recognizes 'honour killings' as a crime in their own right — punishable by death or life imprisonment. The killings, in which families in some rural areas punish women with death if, for example, they marry without permission or have premarital sex, results in about 1,000 deaths every year.

The statistics for girls in schools and universities are also improving. In 1991, women accounted for one in five of the 60,000 students in the country's public universities. By 2001, their numbers had reached a third of the student population of 118,000.

Pakistan's overall efforts in science are on the increase, too. The country's military dictatorships have historically been more generous to scientists than civilian administrations have been, and the Musharraf regime is no different. Science spending has multiplied 60-fold since 1999 and a major programme is under way to upgrade laboratories and boost the numbers of PhDs (see *Nature* 432, 273–274; 2004).

The results of these changes can be seen at the universities in Pakistan's larger cities. At the Usman Institute of Technology in Karachi, a dozen researchers — male and female, young and old — gather to discuss these developments, and how women can make their voices heard in science. Anjum attends, dressed in what has become the international



Hidden talents: chemistry students in Peshawar are among the rising numbers of women studying science.

uniform of Muslim feminism: a long, loose-fitting coat and *hijab*, the scarf that covers hair and head. But she is silent at the beginning of the two-hour meeting. Girls from most Pakistani households are mostly not encouraged to voice opinions in public; in addition, a tradition of deference and respect for older people means that they are unlikely to disagree with the views of senior colleagues present around the table. The ice only really begins to melt when the group gets on to the question of women needing the permission of a male relative to stay late in the lab. Most of the women

acknowledge that they cannot work late without the consent of their parents, husbands or in-laws. The men nod in agreement.

Solo projects

Such attitudes present a significant obstacle to researchers working in rural or remote parts of the country. These places may provide a gold-mine of unexplored opportunities for research in areas such as geology and geophysics, hydrology and biological diversity. But university research managers are wary of assigning such projects to women.

For this reason, says Nayyer Alam Zaigham, a geologist at the University of Karachi, his department has produced just one female PhD in seismology in the past 50

years. "In North America or Europe, you may be able to travel alone," he says. "But in my country, women have to be accompanied. It's a social thing. I cannot send women out to remote areas on their own. I cannot always go with them. And I cannot always arrange an escort. This is a big problem."

Quddusi Kazmi, director of the Marine Reference Collection at the University of Karachi, has plenty of experience of men telling her what she can and cannot do. When she was a student contemplating a research career in marine zoology, she says, one of her professors told her: "We can't allow a lady to do marine zoology. You have to go to the field. You have to go to the sea at odd times. Will your parents allow this?"

Kazmi got her degree anyway, although she says that she will never be allowed to learn how to swim or dive to assist her research. Bathing in public is considered immodest for women, and publicly funded women-only swimming does not exist anywhere in Karachi, she says. While she was doing her degree, collecting specimens meant wading out to sea at low tide and rushing back to shore before the tide came in. "We would feel awkward," she recalls. "People would stare at us as if we were animals." These days, she has to hire fieldmen to do the actual collection.

The good news is that the percentage of female university lecturers has been steadily rising — from 13% in 1990 to 24% in 2002. The female student population of the two largest universities is high: 67% in Karachi and 50% at the University of the Punjab. Even at universities in the more rural provinces of Sindh and Balochistan, one third of the students are women.

It will be tough to get women into an equal position in the lab, admits Rahman. From their earliest years, men and women are discouraged from mixing, he notes. Even in public places, men and women automatically line up in separate queues, and public transport tends to be segregated. But Rahman is convinced that mixed labs are better for research. Having women work alongside men, he says, leads to obvious improvements in the atmosphere of a lab or research institution. "It does a lot of good. We find that men behave better, dress better and are more punctual," he says.

Rahman's director of public relations, Mamoon Amjed, who is sitting in on the interview, breaks into a laugh when she hears this and asks jokingly whether the comment applies to the minister himself. The fact that a Pakistani female civil servant (admittedly, a senior one) can tease a minister in the presence of a journalist is a small sign, perhaps, that life for tomorrow's women may be easier than for those of yesterday. ■

Ehsan Masood is a freelance journalist based in London.



Skeleton keys

Mexican scientists now have the skills and technology to study their backlog of ancient bones. As this treasure trove begins to yield its secrets, Rex Dalton finds local scientists hoping to unravel the mysteries of the earliest settlers of America.

Deep in the jungle of Mexico's Yucatán peninsula, away from the flocks of tourists at the Mayan pyramids and temples, a water-filled cavern is giving local scientists a glimpse much further back in time.

Within this vine-draped sinkhole, a long passageway leads to a near-complete skeleton. Found by archaeological divers a few years ago, early evidence indicates that the remains are at least 11,600 years old — dating back to many millennia before Latin America's great tribal empires. If this age is confirmed, it will be the oldest directly dated skeleton found in the Americas.

"When I was diving and first saw it, I couldn't believe it," says Arturo González González, general director of the Museum of the Desert in Saltillo, Mexico. "I felt sure the bones were very old."

The cave skeleton may provide researchers with an insight into the origins of native American peoples, but it also symbolizes a new era in Mexican palaeoanthropology. For more than a century, the jungle-shrouded edifices of the Maya, Aztec or Olmec peoples have been the primary focus of anthropological research in Mexico. Probing these temples and tombs, some of which date back more than 3,000 years, is a scientific cottage industry and regularly sees foreign scientists troop to the various sites like tourists.

But in recent years, a growing group of Mexican scientists has begun to focus on a much earlier period — when the land now called Mexico was the crossroads for the first inhabitants of the Americas as they traversed the continents. These scientists are only now

scrutinizing Mexico's vast and largely unstudied collection of skeletons, which could be anything from 100 years to tens of thousands of years old. And they are exploring new sites, some of them underwater, from Yucatán to the tip of the Baja California peninsula. The growth in local knowledge, expertise and resources is helping to solve some mysteries — and to uncover new ones — about who populated the Americas, and when.

Out of Asia

The prevailing theory is that most modern native American groups are descended from several waves of people who migrated from northern Asia between 14,000 and 10,000 years ago, crossing into America over a land bridge at the Bering Strait. Although the details of these migrations are not fully understood, it is known that a group called the Clovis people had made it to the southwestern United States by 11,500 years ago, travelling south as the ice that had gripped the north American continent slowly receded at the end of an ice age.

But this cannot be the whole story. Archaeological remains indicate that there were people living in Monte Verde, Chile, at least 12,500 years ago — thousands of years before anyone could have travelled down through the middle of an ice-bound conti-

"Unlike the United States, Mexico sees very little conflict between researchers and indigenous peoples when it comes to research on human remains."

C. THEOBALD/LIVERPOOL JOHN MOORES UNIV.



Remains of the day: using Mexico's vast collection of skeletons (far left) or new discoveries underwater (above), researchers such as Silvia Gonzalez (left) are piecing together the country's ancient history.

ment. Some of those who came over the Bering Strait may have travelled down the coast to reach the south more rapidly, and researchers continue to look for evidence of this coastal migration¹. Another wave of people may have arrived at California directly from southern Asia and the Pacific islands by boat, and worked their way south from there. Some remains have hinted that modern groups of native Americans may have descended from these people as well as from the northern Asian immigrants^{2,3}.

But finding the evidence to support these theories isn't easy. In South America, archaeologists have mainly found signs of habitation, such as ancient campfires, rather than skeletons, making it impossible to work out where these early people came from. And in North America, most if not all of the skeletons studied seem to come from northern Asia — as would be expected, as Pacific Rim immigrants are not likely to have wandered north into chillier climes from their landing sites. Sandwiched between the two continents, Mexico may hold the much-needed evidence to clarify the origins of America's first inhabitants.

Body of evidence

Since the nineteenth century, Mexico has been collecting and storing skeletons of possible interest at the National Museum of Anthropology in Mexico City. This collection has quietly grown in the basement to some 25,000 specimens. Among the skeletons are dozens of specimens that

could open new chapters about the peopling of the Americas. But these bones have so far languished in cardboard boxes, waiting for the day that local scientists have the training, access to technology and money to conduct proper research on them.

To a great extent, that time has now arrived. Mexico's steady scientific development over recent years has given birth to a generation of researchers ready, willing and able to tackle those collections. But in many cases, they are still sorely lacking funds. Richer nations have government agencies to fund such research projects. And for more than a century, wealthy US or European philanthropists have paid for digs at home and abroad. Researchers in a developing nation such as Mexico don't have many governmental or philanthropic options; they must scramble for every peso.

Some local researchers have looked for funding from media outlets, be it *National Geographic* or the Discovery Channel. Others — including González — have had to resort to promoting off-road vehicles to win funds for their work. In exchange for cash, González puts pictures of Dodge trucks in his slide presentations at conferences.

They may have their money worries, but there is one thing González and his colleagues don't have to cope with. Unlike the United States, Mexico sees very little

conflict between researchers and indigenous peoples when it comes to research on human remains. Native American tribes have delayed or halted US genetic studies of early skeletons — including the famous Kennewick man, a 9,000-year-old human skeleton found in Washington state in the 1990s — as they say that such studies would violate their notions of respect for the dead.

In Mexico, the museums and their operations are seen as encompassing all peoples, whether they are indigenous or progeny of natives and Spanish conquerors. Today's population, mainly of mixed origins, has no strong feelings about the treatment of these bones. For the moment, this relatively relaxed attitude is encouraging US researchers to head south.

Ancestral ties

These growing links are also seeing more Mexican researchers collaborate with teams outside the country, which in turn is helping to bolster the nation's skills base. Last November, Alejandro Terrazas Mata, a physical anthropologist at the National Autonomous University of Mexico (UNAM) in Mexico City, began a winter field season in Ethiopia — home to some of the most important specimens showing man's evolution from primate and journey out of Africa. There he worked with a team headed by Tim White, a palaeoanthropologist from the University of California, Berkeley. "We are hoping to transplant some of the remote-sensing and fieldwork methods we've used successfully in Africa to assist Mexican palaeoanthropologists," says White. For Terrazas, it was an opportunity to learn how to do fieldwork in the hotbed of anthropology. "This is a dream of a lifetime," he says.

Silvia Gonzalez (no relation to Arturo), a geoarchaeologist trained in Mexico City and Germany and now at Liverpool John Moores

"We are hoping to transplant some of the remote-sensing and fieldwork methods we've used successfully in Africa to assist Mexican palaeoanthropologists."

— Tim White

University in Britain, has also found useful collaborations. Although a rising star in her field back home, she has had to scramble for funds. Fortunately, she managed to secure sufficient cash in Britain to continue her work, and now has a \$60,000 grant from

Britain's Natural Environment Research Council to study specimens in her homeland, with an emphasis on the interaction between ancient peoples and climate. Although the grant is modest, Gonzalez says that it has attracted collaborations with colleagues from the University of Oxford and the University of Bristol, along with countrymen from Mexico's National Institute of Anthropology and History in Mexico City and other institutions.

One of the most exciting studies in Mexico's anthropology labs today involves

the comparison of ancient and modern bones with those of the Pericú — a tribe that lived near the tip of Baja California from about 2,500 years ago until the late 1800s. The Pericú lived in the hot, harsh desert of Baja in apparent isolation from other humans, surviving on sea life and cacti. The tribe was only discovered by researchers in the late 1940s, and has been a subject of much interest ever since.

Last year, a team of Argentine, Spanish and Mexican researchers published a study of the shape of 33 Pericú skulls found in museum collections^{2,3}. The skulls are long and narrow — similar to skulls in south Asia and the Pacific Rim, and not much like the more rounded skulls typical in northern Asia. This implied that some modern peoples may have evolved from an early wave of Australasian migrants, whereas many had assumed that all descendants had been from the later Asian migrations.

Pacific links

This theory captured the imagination of scientists and the public alike, with press headlines proclaiming that Aborigines from islands off southeastern Asia had founded America. Interest was further inflamed when the media ran a story saying that preliminary DNA tests showed that the Pericú were related to the Maori, a tribe from the Pacific islands. But, at a symposium in Mexico City last September, Phillip Endicott, a PhD student at the Henry Wellcome Ancient Biomolecules Centre at the University of Oxford, UK, revealed that the Pericú DNA matched modern native Americans of north Asian descent. A lone DNA sample reflecting Maori genes could not be replicated, he said, indicating it was probably a contamination from other tests he was performing on Maori samples in the same lab.

Although the team's early expectations of a link between the Pericú and the Maori were dashed, the DNA results posed an even more interesting question: if they're not genetically different, why do the Pericú have skull shapes so different from those of other natives of North America? The Mexican researchers and their colleagues are now exploring a promising hypothesis: over generations, the skulls of the Pericú may have become elongated because the tribe used their teeth as tools for net or fishing-line work. Graduate students at the Autonomous University of Baja California Sur in La Paz



Picking up the pieces: human remains from the early inhabitants of the Americas are still being uncovered in rock shelters in Baja California Sur.

are now examining the teeth closely for signs of wear in an attempt to help corroborate this theory.

But if this idea gains support, it will throw up a fresh problem. For the Pericú it will mean that DNA tests now contradict a study based on skull shape that made it into the pages of *Nature*³. "This could have very profound implications," says Silvia Gonzalez, as it would cast doubt on all past and future work using skull shape alone to make decisions about a skeleton's ancestry. Without supporting evidence from DNA, such decisions could well be wrong, she says.

"To draw any firm conclusions about possible Australasian immigrants to early America, far more bones will need to be analysed."

To draw any firm conclusions about possible Australasian immigrants to early America, far more bones will need to be analysed. So that is what Silvia Gonzalez's group is doing now. Team-member Alfonso Rosales López and his colleagues at the National Institute of Anthropology and History Museum in La Paz have been collecting Pericú specimens from rock shelters, cave floors or coastal shell middens for more than a dozen years. The collection is so substantial that even some of the more extraordinary finds — including a skull where the back half was cleanly cut off by a blow from something like a machete — have not yet been scientifically described. The skull slice is now being studied as possible evidence of cannibalism or

some kind of ritual. Although Rosales López is a co-author of a book on the Pericú culture⁴, his publications have been limited because of monetary and language barriers. That is changing now, thanks to Gonzalez's funding and her help with translation.

Meanwhile, researchers are hot on the trail of much older skeletons to investigate similar questions. Among the most anticipated results is evidence from DNA extracted from 'Peñon woman'. Found in 1959 near the airport outside Mexico City, this skull has been dated to 10,755 years ago⁵. And its DNA is the oldest sample yet extracted from a skeleton in the Americas. Endicott is now analysing DNA from Peñon woman. He declined to discuss his results until they are published or presented at a conference.

Arturo González is also on the lookout for more specimens from the watery caves of the Yucatán, searching for ancient bones that may yield DNA. Many of the bones he has studied so far were found by recreational divers, who spotted them while exploring the

kilometre-long networks of caverns. Amid the remains of camels and elephants that died off during an ice age some 12,000 years ago, they have found some of America's earliest inhabitants. González is slowly retrieving these bones for study, and sharing the results with local colleagues at conferences such as last September's symposium in Mexico City.

Determining the age of skeletons after millennia under water "is a mess", acknowledges Ervin Taylor, an expert in dating techniques at the University of California, Riverside. But DNA left in the watery bones of the skeleton from Yucatán so far hints at a date at least 11,600 years old, he notes — if not older. To be sure, "I need more bone", Taylor says.

Arturo González is ready to secure those specimens. He is heading into the caves this spring to follow up new hints of specimens, bones that he hopes will provide sufficient DNA for this new chapter in genetic testing. Together, Mexican scientists are at last pulling skeletons from cardboard boxes, the dusty recesses of museums and watery graves into the light of scientific study. ■

Rex Dalton is *Nature's* US West Coast correspondent.

1. Dalton, R. *Nature* **422**, 10–12 (2003).
2. Dillehay, T. D. *Nature* **425**, 23–24 (2003).
3. González-José, R. et al. *Nature* **425**, 62–65 (2003).
4. Rosales López, A., Fujita, H. *La Antigua California Prehispánica: La Vida Costera en El Conchalito* (National Institute of Anthropology and History, Mexico City, 2000).
5. Gonzalez, S. et al. *J. Hum. Evol.* **44**, 379–387 (2003).

Rebuilding fisheries will add to Asia's problems

Overfishing has already caused depletion and conflict. Instead, train people for new jobs.

Sir— The tsunami that hit south and southeast Asia, taking a horrific toll in human lives, also affected several coastal industries, including tourism, agriculture and shrimp farming, though to what extent is unclear. As noted in your News story “Scientists seek action to fix Asia’s ravaged ecosystems” (*Nature* **433**, 94; 2005), the effects in some areas were exacerbated by existing environmental problems stemming from settlement and industry.

The governments of Thailand and Indonesia have announced some estimates of fishing boats lost and highlighted the need for investments to restart the fisheries. However good their intentions, I believe that Western aid agencies, and indeed, the governments of the region would be ill-advised to rebuild the fisheries as they were before the tsunami.

Apart from oceanic fisheries for tuna and other large fish, fisheries in the tsunami-affected region fall into two categories: ‘artisanal fisheries’, relying on small (5 m or less), owner- or family-operated craft, some unmotorized; and ‘industrial fisheries’, using larger vessels, mainly trawlers but also other

specialized craft with salaried crews.

Jointly, their fishing activities have radically depleted the nearshore resources, down to depths of 100 m in places. Governments in the region have tried to encourage the industrial fisheries to operate farther offshore, but with little success, mainly because biological production, in tropical waters, is much higher inshore than offshore. Hence the artisanal and industrial fisheries essentially target the same shrimp and fish stocks, leading to intense competition.

This competition and the ensuing violence, including boat burnings and riots, can be serious enough to prompt governments to take action, such as the 1980 ban on bottom trawlers in western Indonesia. Usually, however, government policies ignore these conflicts. Sometimes they exacerbate conflict by subsidizing the construction and operation of industrial vessels, even in cases where these do not add to the total catch, but reduce that of the artisanal fisheries.

International aid has often aggravated this through technological and capital transfers, or donations of surplus vessels.

Meanwhile, failed agricultural and social policies aggravate the situation by driving thousands of landless farmers to coastlines, where they usually fail to emulate the more sustainable ways of ‘traditional’ fishers.

After the tsunami, the initial push will be to get people back to the jobs they know, and it will be hard to argue otherwise in the midst of the chaos. But rebuilding the fisheries without structural reform will only intensify these trends and conflicts.

The challenge is to rebuild fisheries while directing as much money and energy as possible to generating land-based job opportunities for young fishers. Emphasis should be given to basic education and technical skills: many fishers in south and southeast Asia are illiterate, and this limits their social mobility.

Amending the old adage that teaching people to fish is better than giving them a fish to eat, we should instead be teaching them to repair bikes, sewing machines and water pumps.

Daniel Pauly

*Fisheries Centre, 2259 Lower Mall,
University of British Columbia,
Vancouver, British Columbia V6T 1Z4, Canada*

Need for a risk-informed tsunami alert system

Sir— The tsunami death toll from the earthquake of 26 December 2004 is a tragedy of such dimensions that it creates a tipping point in hazard perception and risk management, with existing procedures being thrown into a new and critical light (“Inadequate warning system left Asia at the mercy of tsunami” *Nature* **433**, 3–6; 2005).

In particular, an urgent review is needed of the process for issuing tsunami warnings. Developed by Earth scientists over four decades, it has insufficient input from risk specialists who understand the difficulties of emergency decision-making under uncertainty.

Tsunamis occur most frequently in the Pacific Ocean, so existing monitoring and warning systems have been developed for those conditions. Because earthquakes that could lead to a tsunami happen about once a year on average, concern over the economic cost of false alarms means that detecting a very large earthquake is not enough — additional data from pressure sensors or tide gauges are needed to confirm that a tsunami has indeed been generated.

Outside the Pacific Ocean, it would be

impractical to insist on certainty before issuing a warning. Without adequate monitoring of seismic activity and sea conditions, such an approach may be ineffective. Furthermore, in developing nations with high population density, the cost–benefit balance between human safety and financial loss is tilted more heavily towards avoidance of mass casualties than in affluent Hawaii or Oregon.

An alternative approach is to establish evidence-based criteria for raising different levels of tsunami alert. This would reduce the chance of complete alert failure, at a cost of some additional false alerts.

An analogy might be made with gale warnings, whereby the possibility of a severe storm is sufficient for seafarers to be alerted some hours in advance, but not forcibly grounded.

A risk-informed warning system, based on global online tsunami databases (see www.ngdc.noaa.gov/seg/hazard/tsu.shtml and tsun.ssc.ru/tsulab/On_line_Cat.htm), and implemented via the Bayesian Belief Network formalism, could help decision-makers incorporate the intrinsic uncertainties in tsunami forecasting.

Belief Networks are probabilistic inference graphs providing a logical basis for reasoning under uncertainty, and aid decision-making by accounting jointly for

uncertainties associated with accumulated experience (including expert judgement) and with processed statistical data. Such a scheme would provide a rational basis for tsunami alerts, and allow for a new low-alert level, akin to a gale warning, that would advise against sea and beach excursions during times of heightened risk after a major earthquake.

For coastal authorities, a low-level false alarm every few years would be a small price to pay for avoiding a catastrophe. Indeed, such warnings may prove cost-effective, if the return of tourism requires reassurance. And as with fire drills, the occasional alert would serve as a useful reminder to the local community of a latent threat.

Gordon Woo*, Willy Aspinall†

**Risk Management Solutions Ltd,
Peninsular House, 30 Monument Street,
London EC3R 8HB, UK*

*†Department of Earth Sciences,
Bristol University,
Wills Memorial Building, Queens Road,
Bristol BS8 1RJ, UK*

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

Part of the culture

Magazines and journals reveal how the Victorians viewed science.

Science in the Nineteenth-Century Periodical: Reading the Magazine of Nature

by Geoffrey Cantor, Gowan Dawson, Graeme Gooday, Richard Noakes, Sally Shuttleworth & Jonathan R. Topham
Cambridge University Press: 2004. 346 pp. £45, \$75

Science in the Nineteenth-Century Periodical: An Electronic Index

www.sciper.org

Culture and Science in the Nineteenth-Century Media

edited by Louise Henson, Geoffrey Cantor, Gowan Dawson, Richard Noakes, Sally Shuttleworth & Jonathan R. Topham
Ashgate: 2004. 322 pp. £47.50, \$84.95

Science Serialized: Representations of the Sciences in Nineteenth-Century Periodicals

edited by Geoffrey Cantor & Sally Shuttleworth
MIT Press: 2004. 376 pp. £25.95, \$40

Aileen Fyfe

In 1829, the historian Thomas Carlyle noted that “every little sect among us, Unitarians, Utilitarians, Anabaptists, Phrenologists, must have its Periodical”. By the end of the century, periodical publishing had expanded dramatically: there were more titles, appearing more frequently, and reaching ever greater audiences. Periodicals were the closest Victorian Britain got to the mass media.

In 1999, the Science in the Nineteenth-Century Periodical (SciPer) project was established at the universities of Leeds and Sheffield to investigate the ways in which the sciences were discussed and represented in this important medium. Scientific journals had already been well studied, thanks to their importance as a means of communicating new research between members of the scientific community. Instead, the SciPer team decided to study generalist publications, ranging from such well known literary journals as *The Cornhill Magazine*, to *The Wesleyan-Methodist Magazine*, the comic journal *Punch* and even *The Boy's Own Paper*. These sorts of publications were widely read both within and outside the scientific community. Studying them is the equivalent for the nineteenth century of thinking about how modern science is portrayed on television and radio, and in the press. Such studies can tell us a vast amount about what non-scientists think of science and its practitioners.

The SciPer project is now coming to a close, and the annotated electronic index of the scientific references, allusions and commentaries in 16 generalist periodicals



A bright new age: an 1889 *Punch* cartoon said women should shade themselves from electric lights.

has just been published. Researchers can not only locate reports of Michael Faraday's lectures, or reviews of Charles Darwin's *On the Origin of Species*, but also uncover references to their favourite scientific topics buried deep within short stories, satirical poems, travelogues and articles on other unrelated areas. As well as the electronic publication, the SciPer team has also published a volume of essays, *Science in the Nineteenth-Century Periodical*, that discusses the representation of the sciences in the 16 periodicals that they have studied in depth over the past six years. They have also edited two volumes of conference proceedings, *Culture and Science in the Nineteenth-Century Media* and *Science Serialized*.

The conference proceedings illustrate Carlyle's point about the sheer variety of nineteenth-century periodicals, and show that science crops up in all sorts of unexpected places. Yet they also demonstrate the current approach of most historians to using periodicals: with some notable exceptions, it tends to be a case of turning to the highbrow reviews in search of reactions to particular scientific books, and using the literary weeklies to track the reporting of scientific meetings and discoveries. It is a bit like trying to write about science in twentieth-century culture by analysing the *London Review of Books* and the output of BBC Radio 4. Thus, *Science in the Nineteenth-Century Periodical* serves to remind us of the existence of many other sorts of periodical. To get a full picture

of the myriad ways in which science appears in the public arena, we would need to look not just at the *London Review of Books* but also at *Good Housekeeping*, *The Tablet* and *Private Eye*.

There is no doubt that the electronic index will be very helpful to all those hoping to unearth intriguing snippets of science in unlikely places, such as descriptions of Charles Babbage's designs for a calculating engine in *The Wesleyan-Methodist Magazine*, instructions on making your own optical toys in *The Boy's Own Paper*, and the intriguing political messages that *Punch* conveyed through an illustration referring to an 1868 solar eclipse in India. Nonetheless, the essays in *Science in the Nineteenth-Century Periodical* — and especially the well written and thought-provoking introduction — urge us to consider the context in which those snippets were found. A report on a new medical breakthrough in *Private Eye* is likely to serve quite a different purpose from a mention of the same development in *Good Housekeeping*. The SciPer essays demonstrate, for instance, how *The Wesleyan-Methodist Magazine* was a different sort of publication from *The Review of Reviews*, and how those differences of editorial policy, intended audience and internal format profoundly affected the ways in which the sciences could be reported, debated or criticized in each periodical.

Those who believe that we now live in a world with 'two cultures' — science and the

arts — often see the origins of this situation in the late nineteenth century. As science became more specialized, its practitioners more professionalized and its language more technical, it is often thought to have become separated from the rest of culture. *Science in the Nineteenth-Century Periodical* shows that this is not an accurate picture of what was going on, and must make us question whether it holds for the twentieth century either. True, there were increasing numbers of specialist science journals, but there were increasing numbers of specialist journals in every other field of endeavour too. It isn't true that the sciences disappeared from generalist publications in the late nineteenth century, nor that the science which did appear there was "mere popular science".

Men of science continued to be part of an intellectual community that debated the issues of the day in the highbrow journals. *Science in the Nineteenth-Century Periodical* reveals that some new areas of science were actually pioneered in generalist periodicals before they were picked up by the scientific community. For example, theories of infant development were being discussed in *The Cornhill Magazine* at a time when most men of science would have scorned to enter the female domain of the nursery.

The SciPer project is not just about the popularization of science. Studying the sciences over the whole range of the nineteenth-century periodical press allows us to witness crucial debates about the formation of disciplines. Sometimes the pages of periodicals reveal constructive interplay between experts in two fields that we would now consider quite remote. At times they reveal non-scientific intellectuals appropriating a scientific theory for an intriguingly different purpose. Occasionally, a new theory or approach is presented that will later be incorporated into science, such as the case of infant development. And sometimes experts or journalists explain science to audiences who might not understand all the technicalities — but that is only one small part of the discussion of science in those periodicals.

The rhetoric of 'popular science' or of the 'public understanding of science' is very much a concern of a world where we can see a clear demarcation between 'science' and 'non-science'. But in the nineteenth century, there was still an enviable mixing and cross-fertilization between seemingly disparate subjects, and there was no clear consensus on what counted as science — let alone on how a 'public' understanding might differ from any other sort of understanding. Consensus was eventually reached, not in the pages of the specialist scientific journals, but in the pages of the generalist periodical press. ■

Aileen Fyfe is at the Department of History, National University of Ireland, Galway, Ireland.

SciPer project

▶ www.sciper.leeds.ac.uk



Enrico Fermi (left) probed the structure of nuclear particles with the Chicago synchrocyclotron.

Fermi's legacy

Fermi Remembered

edited by James W. Cronin

University of Chicago Press: 2004. 296 pp. \$45

Enrico Fermi: His Work and Legacy

edited by Carlo Bernardini & Luisa Bonolis

Springer: 2004. 380 pp. £30.50, \$49.95, €39.95

Giulio Maltese

Enrico Fermi was one of the most influential physicists of the twentieth century. He was born in Rome in 1901, and became Italy's first professor of theoretical physics, aged just 25, at the University of Rome. He created some of the cornerstones of modern theoretical physics, such as the Fermi-Dirac statistics, which explain the quantum behaviour of electrons, protons and neutrons, and the theory of beta decay, a radioactive process in which certain nuclei emit electrons or positrons. He also discovered the artificial radioactivity caused by neutron bombardment, and was awarded a Nobel prize in 1938 for his discovery of the properties of slow neutrons.

That same year he fled Italy to escape the racial laws, which affected his wife Laura, and the poor research conditions. He accepted an appointment at Columbia University in New York, where he continued his research on neutrons. In 1942 he moved to the University of Chicago, where he achieved the

first self-sustained nuclear reaction. Two years later he went to Los Alamos to help construct the first nuclear bomb.

On his return to Chicago after the war, he founded the most important school of physics in the world: many of its students were to be outstanding figures in physics in the second half of the century. He also came back to theoretical and experimental physics, achieving important feats in both fields, such as the idea of the compound pion — that pions are composed of a nucleon and an antinucleon — and the hint of the first pion-nucleon resonance, obtained with the Chicago synchrocyclotron, which came into operation in September 1951.

In 2001, to commemorate the centenary of his birth, celebrations were held in both Italy and Chicago. The University of Chicago organized a symposium, and this led to the book *Fermi Remembered*. The book contains recollections from some of Fermi's colleagues, a description of Fermi's impact on physics, and some well selected material from the Fermi archives in Chicago. It opens with a biographical sketch by Fermi's former student Emilio Segrè, first published in Fermi's *Collected Papers*. Also intriguing is the correspondence between Fermi and Leo Szilard in 1939 describing the genesis of some fundamental ideas about the chain reaction. Fermi's own account of those years is reprinted from *Collected Papers*. The book also includes letters and documents from the postwar years concerning scientific, political and personal topics, giving insights into Fermi's character and into his scientific

Science in culture

A galaxy of elements

It's still the periodic table — but with a twist.

Martin Kemp

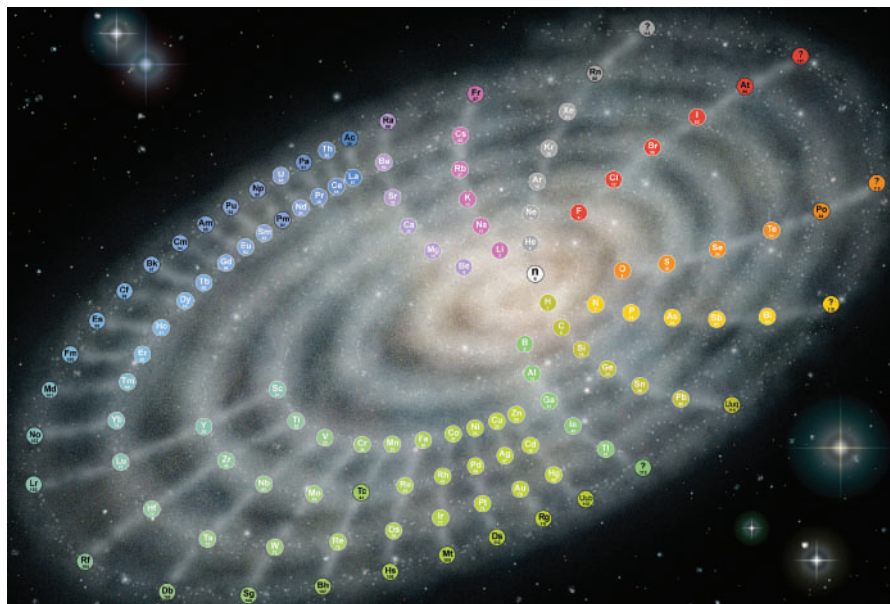
When Dmitri Mendeleev unveiled his periodic table of the elements in 1869, his motives in part embraced what can broadly be described as aesthetic impulses. He sought to define the elusive “unity of matter” through the blending of experimental rigour with pythagorean harmonics (see *Nature* **393**, 527; 1998). If we recognize that a fundamental instinct for the beauty of implicit order has fired the endeavours of pioneering scientists over the ages, we should also acknowledge the role of aesthetically engaging images in drawing non-professionals into the wonder of scientific understanding.

The new ‘galactic’ version of the periodical table, devised by Philip Stewart of the Department of Plant Sciences at the University of Oxford, UK, is designed to achieve precisely this latter goal. Stewart is by no means the first to transform Mendeleev’s rectangular table into a spiral: his scheme stands in a long tradition of alternative configurations, both flat and three-dimensional.

The primary order of the elements is their linear array according to increasing atomic weight (or the positive charge of the nucleus). The secondary relationships in the various schemes have been paraded in various ways, according to the different priorities, interests, needs and intuitions of those who have devised them. Their graphic rendering reflects, with varying degrees of success, the way their designers envisaged the audience: young or old, naive or knowledgeable.

As a boy, Stewart was inspired by a mural at the Exhibition of Science in South Kensington, London, during the Festival of Britain in 1951 — as indeed were others who were later to enter the world of science. The mural, produced by the artist Edgar Longman, depicted the periodic table as a multicoloured, elliptical spiral of box-like sections.

The ellipse as an iconic configuration in modern science has its own history. Before Johannes Kepler’s definition in the early seventeenth century



of the elliptical tracks of the planets, the circle or sphere was the shape that ruled supreme in defining the most perfect of cosmic orders. It is interesting (and unexplained) that, at the same time, ellipses and ovals gained prominence in Baroque ecclesiastic architecture in place of the circles, crosses and polygons favoured in the Renaissance.

Modern astronomy has ensured that we now readily recognize the thrilling dynamism of galaxies in any elliptical array of bright bodies. Stewart’s new periodic table links the elements in their primary sequence with the dust of multitudinous stars. The spokes, curved by the pull of a notional attractor to the upper right, are composed of wisps of interstellar cloud. Such overtly starry allusions stand in a subtle balance with the genuine chemical advantages in the relative positioning of the elements and their associations. For example, hydrogen sits “comfortably” above carbon, Stewart explains. Lutetium and lawrencium,

which cause problems for the conventional periodic table, can be seen here both as the last of the lanthanides and as the first of the next block of transition metals. The placing of neutronium, ‘element 0’, at the very heart of the galaxy is particularly elegant.

At a time when stunning images of Titan, Saturn’s largest moon, from the probe Huygens are being shown on our television screens, with the black-and-white images artificially rendered in colour, it is appropriate to acknowledge the key role played by beauty in engaging a wide range of spectators with science. Engagement is a necessary prelude to communication. As Stewart says: “Science needs the emotions as well the intellect. Young people must have enthusiasm to sustain them in the study of difficult subjects.”

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK, and co-director of Wallace Kemp Artakt.

♦ www.chemicalgalaxy.co.uk

and teaching activities from 1945 to 1954.

There are several articles written by Fermi’s research colleagues and by students “who were in that magic environment at the Institute for Nuclear Studies during the Fermi years”. The book ends with an essay by James Cronin evaluating the predictions on the future of particle physics that Fermi made in a speech in 1954 to mark his retirement as president of the American Physical Society. This book will interest both specialist and general readers, as it provides valuable archive material and sketches of Fermi’s life, as well as personal reminiscences from his former students and collaborators.

As part of the Italian celebrations of

Fermi’s centenary, a book was produced as a resource for physics teachers in secondary schools to introduce their students to Fermi’s science. *Enrico Fermi: His Work and Legacy* is an English translation of this book. It includes commemorative essays by Edoardo Amaldi, one of Fermi’s Italian disciples, and by Fermi’s colleagues and friends Enrico Persico and Franco Rasetti. There are a further 14 essays on several fields of physics that benefited from Fermi’s work, including statistical mechanics, quantum electrodynamics, non-linear systems and particle physics.

The book concludes with a chronological essay by Luisa Bonolis on Fermi’s work. Some articles, such as those on the development of

nuclear physics, on Fermi’s legacy in particle physics, on weak interactions and on non-linear systems, provide interesting accounts of subsequent developments in these fields. These reveal the importance of Fermi’s contributions in preparing the ground for future work.

It is not only physics teachers who will enjoy this book, but also physicists and science historians who want to know more about why Fermi’s work is at the core of so many modern views of physics.

Giulio Maltese is a physicist and a historian of physics based in Rome. He wrote *Enrico Fermi in America: A Scientific Biography* (Zanichelli, 2003).

Life and death in Sumatra

Among Orangutans: Red Apes and the Rise of Human Culture

by Carel van Schaik, with photographs by Perry van Duijnhoven
Belknap Press: 2004. 272 pp. \$29.95, £19.95, €27.70

Alison Jolly

Aceh in Sumatra was the province nearest the epicentre of the earthquake that struck on 26 December. More than 100,000 people there were killed by the resulting tsunami.

Thousands of others have died in Aceh's 25-year struggle for independence from Indonesia. One of these, called Yus, was the local head of a project to study a unique population of orangutans that gave us a new perspective into the evolution of human culture. It may seem insensitive to ask about the welfare of animals amid the human tragedy, but they too will suffer, not directly from the tsunami but from the chaos to come.

The orangutans of Aceh's lowland swamps live south of the devastated town of Meulaboh. Logging and clearing the forest profits the local people but not the orangutans. Foreign scientists have been banned from the area since 1999 because of civil strife. Carel van Schaik's field station at Suaq Balimbing was burned down by the Indonesian army, and Yus, who tried to keep it going, was killed by rebels or villagers. *Among Orangutans* is van Schaik's tribute to lost colleagues and the study animals he can no longer see.

The Suaq swamp forest is misery for terrestrial humans but paradise for arboreal apes. Its rich food supply means that its red apes spend more than 40% of their time associating with others, unlike solitary orangutans elsewhere in Borneo and Sumatra. Orangutans have always been a paradox: they are champion tool-users in captivity but rarely use them in the wild. The Suaq apes are the only wild orangutans known to make and use tools, and van Schaik argues that it is social living that has allowed them to do it. One kind of tool is a short, peeled stick, which they use to pry out the fat-rich seeds of the cemengang fruit from a case defended by viciously stinging hairs. In neighbouring areas with the same fruits, the apes' lower population density means that special knowledge cannot reliably spread through a population, only from mother to infant.

Studies of the Suaq orangutans tell us that a capacity for socially transmitted culture was probably present in the ancestor of all great apes before the orangutan lineage split from the gorilla-chimp-human lineage about 13 million years ago. They also suggest that orangutans were more social when they



Hanging on: Aceh's orangutans survive in the face of adversity.

first evolved than most are today. Modern orangutan density and sociability have been limited by humans, who have hunted and eaten orangutans since our species arrived on Indonesia some 40,000 years ago.

Even so, the sociability of orangutans is unique among great apes. Females give birth at 8-year intervals, even in Suaq, and do not mature until they are about 15. Males can live until they are 60 at least. This slow-motion life cycle, relative to other apes, means that a growing youngster has ample time to learn the intricacies of the forest, including the whereabouts of hundreds of individual fruit trees and the complex routes required to reach them.

A young male faces a huge life choice. He may remain as a 'subadult' for 25 years after sexual maturity, associating with other males while he sires a few infants by raping unwilling females. When he does commit to maturity, he is in a make-or-break situation. Only one big male, with huge cheek flanges and throat sack, is king of a local area of forest. His thunderous long-call warns away rivals. The local king sires almost all the infants: females come to him during oestrus. Other big males who do not successfully challenge him have no hope of further offspring. Van Schaik speculates that the sociable males of chimpanzees and humans resemble immature orangutan males, rather than the overdeveloped state of orangutan maturity.

All this is explained in clear, readable text, with beautiful photographs by Perry van Duijnhoven. Van Schaik gives just one final chapter to the Aceh conflict, the destruction of his study camp, the murder of Yus, and the bleak prospects for the Suaq swamp — without, of course, any idea of the tsunami that

was to wipe out the surrounding towns, without apparently stopping the civil conflict.

A similar book published a few years ago, *In the Kingdom of Gorillas* (Simon & Schuster, 2001), shows how conservation efforts can survive catastrophe. It is the riveting personal account of Bill Weber and Amy Vedder, who spent 20 years trying to establish viable conservation in Rwanda. They joined Dian Fossey's camp at Karisoke in 1978, when poachers were killing an average of ten gorillas each year. Fossey was by then a paranoid recluse who tormented her students and rarely visited the gorillas she loved. Weber and Vedder are unsparing in their portrait of her, but recognize that she was a tragic hero in the Greek sense: her single-minded obsession was both her greatness and her destroying flaw.

The two students were horrified by the shambolic field site, run with no sensitivity for Rwandan needs or Rwandan politics, and no vision of a future different from the past. The World Bank was proposing to clear a third of the Virunga National Park's forests to raise cattle (40% had already been felled in 1959). Dian was paralysed, saying only: "The gorillas will all die." The two young students fought back and eventually the scheme was dropped. Weber and Vedder then spent decades helping to develop a Rwandan constituency for the national parks by local education, by using inspired Rwandan staff, and above all by establishing ecotourism. Before the 1994 genocide, in which some 800,000 Rwandans died, tourism to see the gorillas brought in a third of Rwanda's foreign exchange — the only reason why the gorillas, unlike so many humans, were spared.

Rwanda was riddled with AIDS, poverty and overpopulation, and was the epitome of hatred between neighbours. Weber and Vedder tell of the dedication of foreign conservationists, and of the heroism of Rwandans who risked their own lives to save others. Theirs is a story of hope: hope that even the worst horrors can be forgiven, and that, even in the worst catastrophes, some space for nature, and human feelings, can survive.

One of these books was written after the catastrophe of Rwanda, and one just before the catastrophe of Aceh. Both are required reading for anyone who wants to understand our nearest relatives, the great apes, in a world full of humans, and who believes that our cousins also deserve to live.

Alison Jolly is in the Department of Biology and Environmental Sciences, University of Sussex, Brighton BN1 9QG, UK.

Schrödinger's mousetrap

Part 3: Negative thoughts.

Liesbeth Venema

"Was there any reason for you to dislike Rufus Jaeger?" The question was meant to provoke, but Lister was impressed with the relaxed manner in which Fenton Baumgarden reacted. He was leaning casually against the coffee machine and didn't shift his position one inch.

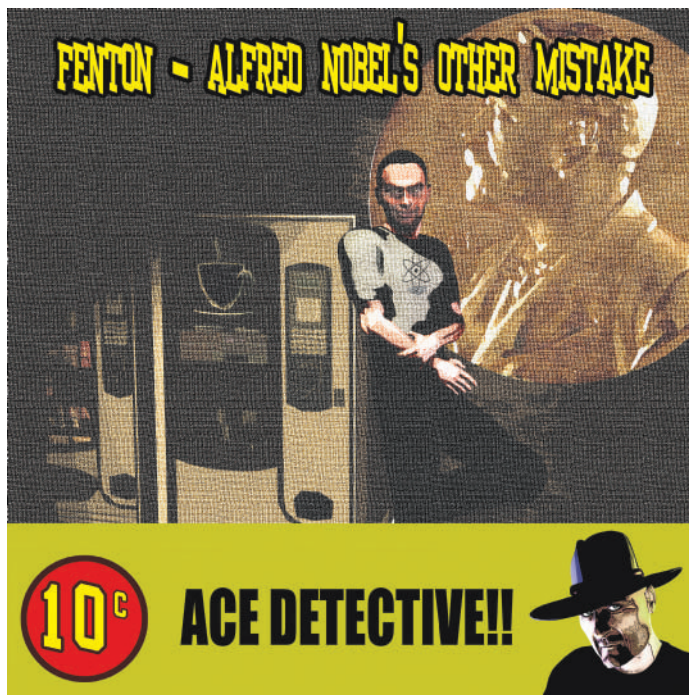
Baumgarden was athletically built, and his skin had been generously tanned by the Californian sun. Although nearing 50, he still got away with wearing a T-shirt depicting a faded logo of a long-forgotten university event, and with sporting a ponytail.

"Sure, quite a good reason I suppose. Not good enough for me though. Life's too short to be bothered by that sort of thing." Baumgarden grabbed the cup that the machine had spewed out and sat down opposite Lister, who fumbled through his notes. "You've probably heard the rumour by now," he continued. "It was Rufus's negative reference that supposedly prevented me getting the Nobel prize. Might or might not be true. To be honest, I think the Nobel is too heavy a burden to carry, I'm really better off without it. It's always a lottery anyway." Baumgarden took a tentative sip from the cup. "Rufus was a brilliant physicist and I respected him for that."

Lister briefly studied Baumgarden, but could see nothing hidden. "Let's get back to recent events," Lister said. "You said you were by yourself in one of the study rooms before the lecture, while everyone else was in the hallway for the coffee break. Are you not interested in socializing?"

Baumgarden gulped the contents of the cup down. "I would've liked to join the others, but I needed to get my head together to act as chair at Rufus's lecture."

Lister nodded absently. "Do you have an idea of what



went wrong during the demonstration?"

Baumgarden moved to the edge of the seat and looked thoughtful for a few moments. "There is something. It came to me almost straight away, but I wanted to think about it." He stood up and paced up and down the room a couple of times and finally went over to the blackboard on the wall, where a schematic of the 'mousetrap' set-up swiftly appeared from his hand. "Laser beam enters this arm here, encounters prism, goes this way. But consider for a moment what will happen if the prism refracts the beam in exactly the opposite direction, which would seem physically impossible. If you work out the optical path anyway, it goes like this, straight to the guy sitting in the chair here."

Lister scribbled quietly in his notebook. "And how can you make the prism refract in the ... er ... opposite direction?"

"That's the thing. As I said, it's not possible. At least, up till now. Negative refraction was discovered in 2001 for microwaves, but so far has not been possible for light. But now there's this new synthetic optical material. Looks just like glass, but refracts light in the opposite, negative direction. Gives totally weird effects. Petra Pruszczyński has called it negadex — it's her group that developed the material. It's still top secret."

"So how do you get to know all about this negadex?" Lister wanted to know.

"Petra has just submitted a paper about the discovery to *Nature* ... and I am one of the

referees." Baumgarden made a dismissive gesture. "Look, it's probably a crazy thought. In any case I can't believe Petra could be involved, she has way too much sense for that. But I had to tell you about the possibility." Baumgarden sat down again and folded his arms. "Anything else I can help you with?"

Lister held up his hand to indicate he needed a moment to arrange his thoughts. So the demonstration set-up was probably compromised during the coffee break, as it now turned out by replacing a prism with a piece of negadex — a material that only a few people in the world knew about. Petra Pruszczyński was present at the conference, and was already on Lister's list of suspects. He needed to find out whether anyone else with ties to her group was present. Baumgarden him-

self couldn't be ruled out. Who else? Lister had a brainwave.

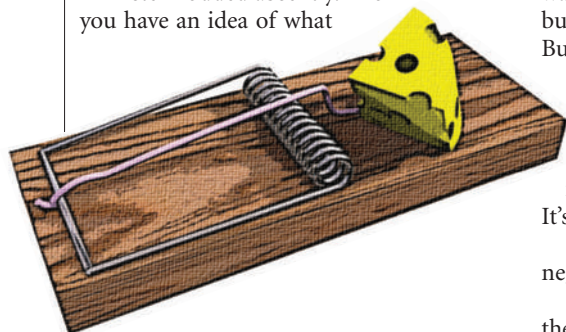
For the first time during the interview, Baumgarden looked a little uneasy. The detective had surprised him by asking which editor was handling Pruszczyński's paper. "Nigel Lorimer. Yes, he is here too..."

Lister sensed that there was more to come. Baumgarden fiddled with his sunglasses. "I should probably tell you before you hear this from someone else. There was a paper from Rufus's group, some five years ago. It was generally seen as a huge breakthrough, and received massive publicity. But it turned out soon after that there was a crucial error in the data that should have been spotted.

"Rufus himself started an internal investigation, no doubt to head off criticism. He cleverly accepted responsibility for the mistake, but made sure it was clear that it was a student who had messed up. Rufus actually managed to emerge with his reputation largely unstained." Baumgarden looked at the floor and shook his head. "The paper had to be retracted. Poor Nigel was devastated — he was the editor who handled it. Nigel is a very decent guy but he tends to take things personally. I think he has been rather obsessed about it ... felt the whole affair damaged him. He tried to keep up appearances and stay polite with Rufus, but I think it must have been clear to everyone that he hated him."

To be continued...

Liesbeth Venema is a senior physics editor at Nature.



Hot pursuit of missing matter

J. Michael Shull

Astronomers are going to extraordinary lengths in the quest to tot up the 'ordinary' matter in the Universe. The latest initiative has probed hot gas in intergalactic space by means of an X-ray lighthouse.

As cosmological measurements have become feasible and more precise, astronomers have undertaken 'cosmic inventories' of mass and energy in all their forms. Nature is quite devious, hiding matter in a range of astronomical objects (stars, planets, galaxies), in intergalactic gas at temperatures of 10^4 K to 10^7 K, and in even more exotic forms — so-called dark matter and dark energy. However, recent inventories^{1,2} suggest that at least 40% of ordinary matter is missing. Using spectrographs aboard the Chandra X-ray Observatory, a group of X-ray astronomers believe they have detected this missing matter, hiding in gas at temperatures of around a million kelvin, and spread throughout vast volumes of intergalactic space (Nicastro *et al.*, page 495 of this issue³).

According to the standard cosmological model, the Universe consists primarily of dark matter (25%) and dark energy (70%), with a smattering (5%) of normal matter known as 'baryons'. To particle physicists, 'baryon' is a collective term for three-quark, subatomic particles (protons, neutrons, and so on). Astronomers would more simply define baryons as ordinary matter (hydrogen, helium and heavier elements) that forms the planets, stars and galaxies, and gas in space. Fortunately, we have a good idea of the total matter to look for. Studies of light elements formed in the Big Bang⁴ and fluctuations in the cosmic microwave background⁵ have derived a precise value for the total baryon density in the Universe: $4.6 \pm 0.2\%$ of the total density of mass/energy.

Although we know how much matter we are looking for, an embarrassing puzzle is that only a fraction of this baryonic matter can be accounted for in known forms^{1,2}. Surprisingly, luminous ordinary matter (stars and galaxies) accounts for less than 10% of the baryons. The remainder must reside in baryonic dark matter, or perhaps it is hiding in some form that is difficult to observe.

Ultraviolet and X-ray spectroscopic measurements suggest that many of the missing baryons lie in low-density gas, distributed

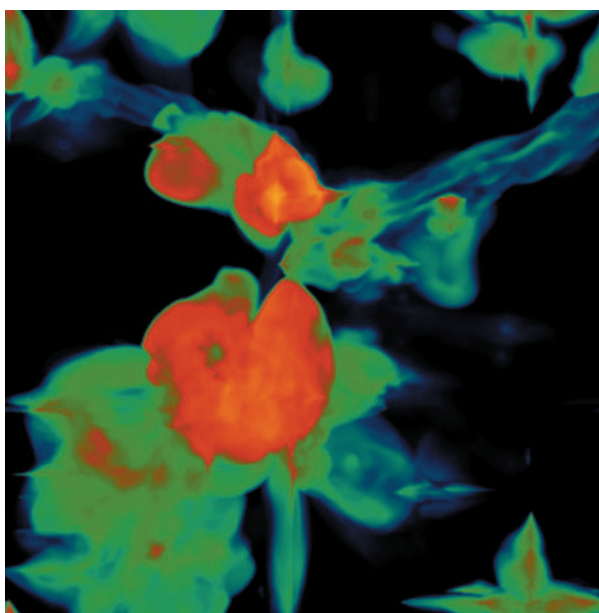


Figure 1 Getting warmer? Galactic winds together with gravitational collapse should have a dramatic effect on the distribution of hot intergalactic medium, as shown by these numerical hydrodynamical simulations; the 'real' image would be about 50 million light years across. Ultraviolet⁷ and X-ray³ observations may have detected the hot gas at temperatures of 10^5 – 10^6 K (red regions) around galaxies and warm-hot intergalactic medium (WHIM) filaments at redshift $z = 0$, corresponding to the local vicinity of the Milky Way and surrounding galaxies. The temperature of the darkest regions is 10^3 K and that of the green regions 10^4 – 10^5 K. (Image courtesy of R. Cen and K. Nagamine.)

throughout the intergalactic medium (IGM), where it has not yet coalesced into galaxies. Although the average density of this IGM is extraordinarily low, 1 million to 10 million times less than the interstellar gas in the Milky Way, intergalactic space is a vast reservoir. Observations of the nearby Universe suggest that only 5–10% of the baryons have collapsed into galaxies and clusters of galaxies. Another 40% can be inferred from absorption lines of cool neutral hydrogen⁶ and in much hotter gas⁷ detected in the ultraviolet absorption lines of O VI (five-times-ionized oxygen). Although these species are trace constituents of the IGM, astrophysical corrections allow one to account for the accompanying amounts of ionized hydrogen.

Nevertheless, these reservoirs of intergalactic gas (H I and O VI) do not solve the puzzle, because 40–45% of the baryonic matter is still missing. Cosmological simulations^{8,9} suggest another hiding place for

the missing matter, because gravity shapes the large-scale temperature structure of the IGM. These models predict that 30–40% of the baryons reside in gas at around 10^6 K, shock-heated by gravitational collapse and galactic outflows driven by supernova explosions (Fig. 1). At these high temperatures, low-density gas is nearly invisible at optical wavelengths. However, heavy elements such as oxygen, nitrogen and neon still retain a few bound electrons. Their ions can be detected by absorption lines in the ultraviolet (O VI) and soft X-ray (O VII and O VIII) bands.

The Chandra and XMM-Newton X-ray observatories offer the chance to probe the hottest portion of this gas, known as the 'warm-hot intergalactic medium', or WHIM^{8,9}. Nicastro *et al.*³ devised a clever experimental strategy to find an 'X-ray lighthouse' to shine through the WHIM. They simply waited for an active galactic nucleus to go into an unusually bright state. When Markarian 421, an exceptionally bright, variable active galaxy, flared twice, they observed it for 200,000 seconds with the Chandra low-energy transmission grating spectrometer. In an exquisite spec-

trum of the hot intergalactic gas observed during this time, they saw the expected absorption lines of highly ionized oxygen and nitrogen. Of special interest were two distant absorption systems, at distances of 150 million and 380 million light years, assuming the usual Hubble expansion law for redshift factors $z = 0.011$ and 0.027 .

These two O VII absorption systems provide tantalizing evidence for the predicted filaments of hot (WHIM) gas at 10^6 K. With only two absorbers along a short path length, it is difficult to know whether this region is typical of the entire Universe. Using admittedly large statistical uncertainties, the authors estimate that this hot IGM could contain the remaining 30–50% of missing baryons, assuming that oxygen has an abundance equal to 10% of that in the Sun. This is a standard assumption for the IGM, but one that needs verification with further work on IGM heavy-element abundances.

What's next in this quest? First, it would be valuable to carry out X-ray observations along other intergalactic sightlines, because Markarian 421 may probe an atypical region of the IGM. Unfortunately, most other quasars are not bright enough to provide such high-quality data. Astronomers have had considerable success probing more than 40 sightlines for O VI absorption lines ($10^{5.5}$ K gas) using the Far Ultraviolet Spectroscopic Explorer (FUSE). But current X-ray and ultraviolet observatories lack the light-gathering power to complete the baryon inventory properly.

That task awaits a new generation of larger space observatories, including NASA's Constellation-X mission¹⁰ and the European Space Agency's XEUS mission for X-ray spectroscopy. Large ultraviolet and optical space telescopes are in the planning stages¹¹, as astronomers seek to replace the spectroscopic capabilities of the Hubble Space Telescope. New ultraviolet and X-ray observatories are needed to complete the

inventory of missing baryons. But they will do much more, allowing astronomers to map out the cosmic web of filamentary intergalactic matter from which the first galaxies and stars were formed. ■

J. Michael Shull is in the Department of Astrophysical and Planetary Sciences, University of Colorado, 391-UCB, Boulder, Colorado 80309, USA.

e-mail: mshull@casa.colorado.edu

1. Fukugita, M. & Peebles, P. J. E. *Astrophys. J.* **616**, 643–668 (2004).
2. Shull, J. M. in *The IGM/Galaxy Connection* (eds Rosenberg, J. L. & Putman, M. E.) Vol. 281, 1–10 (Kluwer, Dordrecht, 2003).
3. Nicastro, F. *et al. Nature* **433**, 495–498 (2005).
4. Kirkman, D., Tytler, D., Suzuki, N., O'Meara, J. M. & Lubin, D. *Astrophys. J. Suppl.* **149**, 1–28 (2003).
5. Spergel, D. N. *et al. Astrophys. J. Suppl.* **148**, 175–194 (2003).
6. Pentton, S. V., Stocke, J. T. & Shull, J. M. *Astrophys. J. Suppl.* **152**, 29–62 (2004).
7. Savage, B. D., Sembach, K. R., Tripp, T. M. & Richter, P. *Astrophys. J.* **564**, 631–649 (2002).
8. Cen, R. & Ostriker, J. P. *Astrophys. J.* **514**, 1–6 (1999).
9. Davé, R. *et al. Astrophys. J.* **552**, 473–483 (2001).
10. NASA, Constellation-X Observatory website <http://constellation.gsfc.nasa.gov>
11. Space Ultraviolet Observatory *Tracing the Cosmic Web* <http://xxx.lanl.gov/abs/astro-ph/9907101>

colleagues⁸ show that the answer lies in differences in the flanking regulatory region of the *yellow* gene. They have found that if this regulatory region from *D. biarmipes* is inserted into the developing wing of *D. melanogaster*, it drives the expression of gene constructs in the distal–anterior region of the wing. This property resides in a discrete regulatory element, known as an enhancer, that drives wing-specific gene expression. The enhancer is also present in the non-spotted species but, in *D. biarmipes*, it contains a short sequence that binds unknown transcription factors responsible for driving distal expression. The enhancer also contains defined sites that bind the transcription factor Engrailed, which restricts expression to the anterior side. Further research may reveal additional modifications in other regulatory regions of *yellow*.

Unfortunately, it is not yet possible to produce genetically modified variants of *D. biarmipes*, so the authors have been unable to carry out the reciprocal experiment: transferring the regulatory regions of *D. melanogaster yellow* into *D. biarmipes*. There may also be relevant differences between the species in the distribution of transcription factors — perhaps explaining why the expression of the *D. biarmipes* gene constructs in *D. melanogaster* is not as sharply defined as that of the normal *yellow* gene in *D. biarmipes*.

Nonetheless, the results so far show that evolution of the dark wing spot in *D. biarmipes* males has involved alterations in an ancestral regulatory element of the *yellow* gene. As a result, the regulatory element has gained multiple binding sites for transcription factors whose distributions define the new region of expression. These distributions, including that of Engrailed, may be deeply conserved features of wing development, and Gompel *et al.*⁸ argue that this process could generate similar spot patterns in distantly related fly species (see, for example, ref. 10). More generally, the evolutionary lability of short protein-binding sequences in the regulatory regions of other key developmental genes might explain the repeated patterns of disappearance and reappearance

Evolutionary developmental biology

How and why to spot fly wings

Paul M. Brakefield and Vernon French

How can different species evolve different physical features despite using similar molecular toolkits? Studies of wing colour development in fruitflies point to specific changes in a gene's regulatory region.

Over the past two decades, comparative studies have shown that the development of even widely disparate organisms uses a surprisingly similar set of mechanisms. The process of development is controlled largely by where and when key genes are switched on or off — events that are in turn determined by the binding of available proteins (transcription factors) to the genes' regulatory regions. From an evolutionary perspective, we need to know which of these developmental components have varied during evolution, providing the raw material that is sieved by natural selection to remould organisms¹. Identifying such components of change requires us to be able to link naturally occurring mutations closely to specific examples of morphological variability, both within species and between related species^{2,3}.

Several studies have begun to reveal that variation in the regulatory regions of particular genes is associated with species differences in insect bristle or pigment patterns^{4–7}. Now, on page 481 of this issue⁸, Gompel *et al.* show how specific changes in the regulatory elements of the gene *yellow* have contributed to the evolution of differences in wing pigmentation among related species of fruitfly.

Yellow is one of the key genes needed to make the melanin pigments in insects. In

fruitflies, it is characteristically expressed in regions that will form darkly pigmented patches on the adult cuticle^{7,9}. In the fruitfly species *Drosophila melanogaster*, *yellow* is expressed at low levels in the wings, which develop a uniformly grey colour. Males of the related species *D. biarmipes*, by contrast, have dark wing tips that are displayed to potential mates during courtship, and the male pupae express *yellow* in matching distal–anterior spots (that is, lying towards the front and tip of the developing wing).

So what determines the difference in the expression of *yellow* between *D. melanogaster* and *D. biarmipes*? Gompel and



Figure 1 Insects displaying the melanin-based colour patterns in their wings. Left, fighting male fruitflies (*Drosophila silvestris*) on Hawaii. Right, mating butterflies (*Ornipholidotos* species) in Kibale Forest, Uganda.

of particular morphological traits in other lineages, such as the horns of dung beetles¹¹.

An intriguing point is that, although Gompel *et al.* show that transfer of the *yellow* gene from *D. biarmipes* to *D. melanogaster* resulted in distal–anterior gene expression, it did not produce a dark wing tip. So, although the regulatory changes that generate a new expression pattern for *yellow* may have been central to the evolution of the *D. biarmipes* wing spot, they are far from the whole story. Expression of the gene *ebony*, for example, is reduced in the spot region of the *D. biarmipes* wing⁹, and when Gompel *et al.* introduced the *D. biarmipes yellow* gene into *D. melanogaster* bearing mutations in *ebony*, a faint wing spot was produced. Clearly, changes in the regulation of *ebony* and other genes involved in the branching pathways of melanin synthesis¹² must contribute to formation of the full wing spot in *D. biarmipes*. By identifying all of these changes and their separate effects, it may be possible to reconstruct the pathway by which the wing spot evolved.

It is not known how the wing patterns of *D. biarmipes* and related species contribute to the performance of these flies in their natural environments. However, behavioural observations on sexual selection in other species, including Hawaiian *Drosophila*¹³ (Fig. 1), encourage the hope not only of identifying the genetic changes that have become fixed during evolution, but also of providing explanations for the morphological changes in terms of natural selection. Related studies of the control of hair (trichome) and bristle development in fruitflies have also begun to reveal some of the underlying regulatory changes in key developmental genes^{4–7}. However, unravelling the adaptive significance of the altered morphologies is likely to prove too great a challenge, as their functions in natural environments remain obscure.

Revealing exactly how diversity in the number, position and shape of patches of wing pigmentation has evolved across the many species of fruitfly will be a challenge for the next decade. However, fruitflies — as well as other insects such as butterflies, with their much richer diversity in wing pigmentation^{3,12} — provide wonderful opportunities for tracing evolution all the way, from specific changes in gene regulation through to the performance of the altered morphologies in nature and to their diversity among related species. ■

Paul M. Brakefield is at the Institute of Biology, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands.

e-mail: brakefield@rulsfb.leidenuniv.nl

Vernon French is at the Division of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK.

e-mail: vernon.french@ed.ac.uk

3. Brakefield, P. M., French, V. & Zwaan, B. J. *Annu. Rev. Ecol. Evol. Syst.* **34**, 633–660 (2003).
4. Sucena, E. & Stern, D. L. *Proc. Natl Acad. Sci. USA* **97**, 4530–4534 (2000).
5. Gompel, N. & Carroll, S. B. *Nature* **424**, 931–935 (2003).
6. Sucena, E., Delon, I., Jones, I., Payre, F. & Stern, D. L. *Nature* **424**, 935–938 (2003).
7. Wittkopp, P. J., Vaccaro, K. & Carroll, S. B. *Curr. Biol.* **12**, 1547–1556 (2002).

8. Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. *Nature* **433**, 481–487 (2005).
9. Wittkopp, P. J., True, J. R. & Carroll, S. B. *Development* **129**, 1849–1858 (2002).
10. Kopp, A. & True, J. R. *Evol. Dev.* **4**, 278–291 (2002).
11. Emlen, D. J., Marangelo, J., Ball, B. & Cunningham, C. W. *Evolution* (in the press).
12. True, J. R. *Trends Ecol. Evol.* **18**, 640–647 (2003).
13. Boake, C. R., DeAngelis, M. P. & Andreadis, D. K. *Proc. Natl Acad. Sci. USA* **94**, 12442–12445 (1997).

Cell biology

Holding sisters for repair

Tatsuya Hirano

When a DNA molecule breaks, its complementary copy can be used as a template for repair. A familiar protein complex is recruited to the damaged site, keeping it close to the undamaged copy.

Before a cell divides, it must replicate its DNA, producing two identical copies of each chromosome, known as sister chromatids. When one of the sister chromatids suffers a break that affects both of its DNA strands, it is mended by 'homologous recombination'. In this process, information on the undamaged sister is used as a template for repair. This mechanism is essential for cell survival and genome stability. But how can the repair machinery find the proper template in the crowded environment of the cell nucleus? Writing in *Molecular Cell*, Ünal *et al.*¹ and Ström *et al.*² provide compelling evidence that a protein complex called cohesin has a crucial role.

The cohesin complex was originally identified as a protein component that ensures the proper segregation of sister chromatids during cell division. It does so by holding them together from the time that they are produced (during the 'S phase' of the cell-division cycle) until the mid-stage of mitosis — the phase in which the sister chromatids of each pair are separated in preparation for making two cells³.

However, previous genetic studies also implicated cohesin in the repair of double-strand breaks (DSBs) between S phase and mitosis (in the post-replicative, or G2, phase). In fact, one of the cohesin subunits had been found to be involved in DSB repair in fission yeast⁴ long before its essential role in chromosome segregation was recognized.

Although subsequent studies in budding yeast⁵ and vertebrates⁶ also supported a requirement for cohesin in post-replicative DSB repair, it remained unclear how the complex participates in this process at a mechanistic level. In principle, two models can be considered. During S phase, cohesin complexes — spaced 10–15 kilobases apart in yeast — build up a physical linkage along the length of the sister chromatids; this genome-wide linkage may be sufficient to keep sisters close enough to allow repair

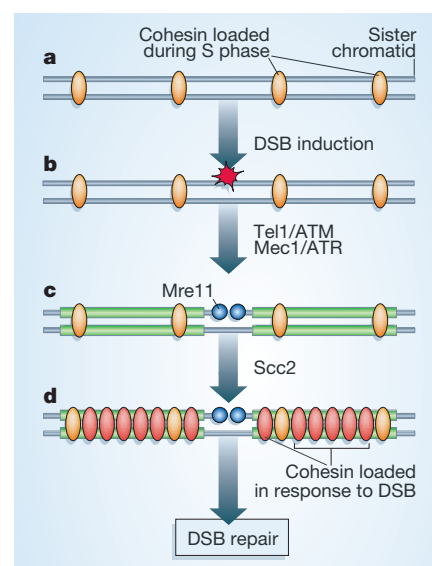


Figure 1 Local recruitment of cohesin for DNA repair. The model shown is based on the findings of Ünal *et al.*¹ and Ström *et al.*². **a**, During the DNA-replication (S) phase of cell division, cohesin (orange ovals) forms links between sister chromatids along their lengths. **b**, **c**, When a double-strand break (DSB) is induced in one of the sister chromatids (**b**), enzymes of the DNA-damage checkpoint — such as Tel1/ATM and Mec1/ATR — are activated. The DSB-repair protein Mre11 (blue circles) binds to the damaged site, and histone protein H2AX is phosphorylated (green bar) in a large chromosomal region surrounding that site (**c**). **d**, The phosphorylation of H2AX allows the *de novo* loading of more cohesin (red ovals), a process that also requires Mre11 and the cohesin-loading factor Scc2. The consequent enhanced linkage between the damaged and undamaged DNA strands allows efficient repair of the DSB.

by homologous recombination when a DSB appears in G2 phase. Alternatively, cohesin may have a more active, more specialized and more local role around the damaged sites.

1. Stern, D. L. *Evolution* **54**, 1079–1091 (2000).

2. Simpson, P. *Nature Rev. Genet.* **3**, 907–917 (2002).

To follow the dynamics of cohesin in response to a DSB in the budding yeast *Saccharomyces cerevisiae*, Ünal *et al.*¹ and Ström *et al.*² have used a well-characterized system in which a DSB can be generated at a defined point of the genome, by inducing the expression of a specific DNA-cutting enzyme called HO endonuclease. By using a high-resolution mapping technique — chromatin immunoprecipitation — both groups find that cohesin accumulates in a large chromosomal region of around 50–100 kilobases surrounding the DSB. Moreover, this local enrichment of cohesin at the DSB site occurs in G2 phase and requires the same protein (Scc2) that facilitates genome-wide loading of cohesin in early S phase. These observations strongly suggest that the DSB induces *de novo* loading of cohesin at the damaged site, rather than a rearrangement of cohesin already loaded on the DNA during S phase.

What is the molecular mechanism underlying this local recruitment of cohesin? A previous study⁷ showed that H2AX — a histone protein, involved in packaging DNA — is phosphorylated in a large region that extends some 50 kilobases either side of an HO-endonuclease-induced DSB in yeast. This modification depends on two enzymes — namely Tel1 and Mec1 (ATM and ATR, respectively, in mammals) — which are involved in a ‘checkpoint’ that delays cell division until damaged DNA is repaired⁷.

Following on from these findings, Ünal *et al.*¹ show that cohesin recruitment in response to a DSB is also regulated by Tel1

and Mec1, and requires H2AX phosphorylation. The formation of the region of phosphorylated H2AX proteins is therefore likely to be a prerequisite for the loading of cohesin (Fig. 1). Equally important, Ström *et al.*² provide evidence that the DSB-induced cohesin loading does indeed establish a *de novo* linkage between the damaged chromatid and its undamaged sister, thereby facilitating DSB repair. Consistent with this conclusion, cohesin is not required for other repair-related processes, such as intrachromosomal gene conversion or the resection of broken DNA strands¹ — events that do not require a template or, therefore, new sister-chromatid linkages.

These observations made in yeast are likely to be highly relevant to our understanding of DSB repair in mammals. In fact, DNA-damage-induced recruitment of cohesin was first hinted at through a lower-resolution, cytological method — immunofluorescent microscopy — in human cells⁸. This recruitment was dependent on the DSB-repair protein Mre11, as has been confirmed in yeast¹. Moreover, other studies showed that ATM phosphorylates one component of cohesin following DNA damage induced by ionizing irradiation, and that this phosphorylation is required for activation of the DNA-damage checkpoint^{9,10}. Cells expressing only a non-phosphorylatable form of cohesin support normal mitotic progression, but exhibit a defective checkpoint response and increased chromosomal aberrations¹¹, implying that this modification of cohesin is specialized for DNA repair in mammalian cells.

In summary, the new studies^{1,2} underscore the importance of a cohesin-mediated sister linkage in DSB repair. They also provide a fresh view of cohesin — which seems to be much more dynamic than had been thought — and raise several new questions. What is the real signal or modification that allows cohesin to accumulate at the DSB site? Does cohesin, or the protein that loads it onto DNA, interact directly with phosphorylated H2AX? What happens when DNA repair is completed? Might cohesin need to be cleaved in order to be unloaded after DNA repair¹², as has been shown in mitosis³? Undoubtedly, answering these questions will further enrich our understanding of the global and local dynamics of chromosome architecture, and its impact on genome stability. ■

Tatsuya Hirano is at the Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, New York 11724, USA.

e-mail: hirano@cshl.edu

1. Ünal, E. *et al.* *Mol. Cell* **16**, 991–1002 (2004).
2. Ström, L., Lindroos, H. B., Shirahige, K. & Sjögren, C. *Mol. Cell* **16**, 1003–1015 (2004).
3. Nasmyth, K. *Annu. Rev. Genet.* **35**, 673–745 (2001).
4. Birkenbihl, R. P. & Subramani, S. *Nucl. Acids Res.* **20**, 6605–6611 (1992).
5. Sjögren, C. & Nasmyth, K. *Curr. Biol.* **11**, 991–995 (2001).
6. Sonoda, E. *et al.* *Dev. Cell* **1**, 759–770 (2001).
7. Shroff, R. *et al.* *Curr. Biol.* **14**, 1703–1711 (2004).
8. Kim, J. S., Krasieva, T. B., LaMorte, V., Taylor, A. M. & Yokomori, K. *J. Biol. Chem.* **277**, 45149–45153 (2002).
9. Kim, S.-T., Xu, B. & Kastan, M. B. *Genes Dev.* **16**, 560–570 (2002).
10. Yazdi, P. T. *et al.* *Genes Dev.* **16**, 571–582 (2002).
11. Kitagawa, R., Bakkenist, C. J., McKinnon, P. J. & Kastan, M. B. *Genes Dev.* **18**, 1423–1438 (2004).
12. Nagao, K., Adachi, Y. & Yanagida, M. *Nature* **430**, 1044–1048 (2004).

Mycology

To be or not to be a lichen

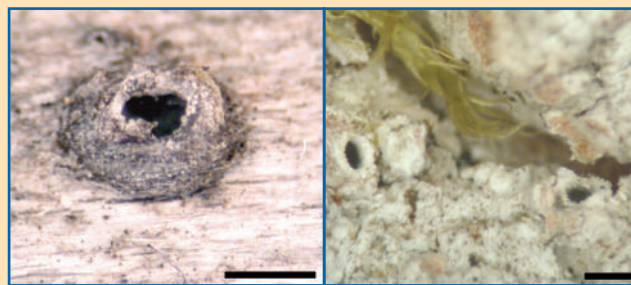
Work by Mats Wedin and colleagues highlights both the versatility of fungi and the complications these unsung organisms pose for the biologist. Two of the ways in which fungi make a living are as saprophytes, drawing sustenance from decaying matter, and as lichens, in which they form an intimate relationship with green algae or photosynthetic bacteria. Evidently, however, a single fungal species can adopt either lifestyle according to circumstance (*New Phytol.* **164**, 459–465; 2004).

Using molecular techniques, Wedin *et al.* looked at fungi living on different parts of aspen (*Populus tremula*) in northern Sweden. On analysing four independent genomic markers, they found that three different species of a fungal genus (*Stictis*) grew directly on wood

without bark as typical saprophytes (left photograph); but when on the bark of the trunk, the same species associated with green algae to form a whitish, crust-like lichen (right). In both images the scale bar is 1 mm.

The lichens had hitherto been placed in a separate genus, *Conotrema*. Thirty-five years ago it was realized that the fruit bodies of *Stictis* and *Conotrema* are microscopically almost indistinguishable, but *Stictis* was not previously recognized as a lichen-forming genus. That species classified in separate genera are in fact the same organism exhibiting different biologies is a startling discovery. In all three cases, the wood-inhabiting non-lichenized and lichenized specimens were mixed together in phylogenetic trees.

Wedin *et al.* point out that the



ability of a fungus to opt for different nutritional modes means that individual species can exploit a wider range of habitats. If the spores that are shot from the fruits of these fungi land on wood, they establish themselves as saprophytes; if they land on bark where appropriate algae are present, they form lichens.

The frequency of this plasticity in lifestyles is unknown. But it may be common, at least in this group of fungi, as the three species able to become optionally lichenized were found on a single species of tree.

The study reinforces the case for never treating lichens as anything other than a lifestyle category among the fungi. It also shows the wisdom of the 1959 decision to restrict the scientific names of lichens to the fungal partners; in consequence, the same names can be used for the fungi regardless of their lifestyle. David Hawksworth

David Hawksworth is in the Facultad de Farmacia, Universidad Complutense de Madrid, Plaza Ramón y Cajal s/n, 28040 Madrid, Spain.
e-mail: myconova@terra.es

Life

In search of the simplest cell

Eörs Szathmáry

Top-down, bottom-up; RNA-based, lipid-based; theory, experiment — there are many different ways of investigating what constitutes a 'minimal cell'. Progress requires finding common themes between them.

In investigating the origin of life and the simplest possible life forms, one needs to enquire about the composition and working of a minimal cell that has some form of metabolism, genetic replication from a template, and boundary (membrane) production. Approaches to this intriguing problem are discussed in Tibor Gánti's *The Principles of Life* (Oxford Univ. Press, 2003), and were also debated at a meeting last December*.

Identifying the necessary and sufficient features of life has a long tradition in theoretical biology. But living systems are products of evolution, and an answer in very general terms, even if possible, is likely to remain purely phenomenological: going deeper into mechanisms means having to account for the organization of various processes, and such organization has been realized in several different ways by evolution. Eukaryotic cells (such as those from which we are made) are much more complicated than prokaryotes (such as bacteria), and eukaryotes harbour organelles that were once free-living bacteria. A further complication is that multicellular organisms consist of building blocks — cells — that are also alive. So aiming for a general model of all kinds of living beings would be fruitless; instead, such models have to be tied to particular levels of biological organization.

Basically, there are two approaches to the 'minimal cell': the top-down and the bottom-up. The top-down approach aims at simplifying existing small organisms, possibly arriving at a minimal genome. Some research to this end takes *Buchnera*, a symbiotic bacterium that lives inside aphids, as a rewarding example (A. Moya, Univ. Valencia). This analysis is complemented by an investigation of the duplication and divergence of genes (A. Lazcano, Univ. Mexico). Remarkably, these approaches converged on the conclusion that genes dealing with RNA biosynthesis are absolutely indispensable in this framework. This may be linked to the idea of life's origins in an 'RNA world', although such an inference is far from immediate.

Top-down approaches seem to point to a minimum genome size of slightly more than 200 genes. Care should be taken, however, in blindly accepting such a figure. For example, although some gene set A and gene set B may

not be common to all bacteria, that does not mean that (A and B) are dispensable. It may well mean that (A or B) is essential, because the cell has to solve a problem by using either A or B. Only experiments can have the final word on these issues.

There was general agreement that a top-down approach will not take us quite to the bottom, to the minimal possible cells in chemical terms. All putative cells, however small, will have a genetic code and a means of transcribing and translating that code. Given the complexity of this system, it is difficult to believe, either logically or historically, that the simplest living chemical system could have had these components.

The bottom-up approach aims at constructing artificial chemical supersystems that could be considered alive. No such experimental system exists yet; at least one component is always missing. Metabolism seems to be the stepchild in the family: what most researchers in the field used to call metabolism is usually a trivial outcome of the fact that both template replication and membrane growth need some material input. This input is usually simplified to a conversion reaction from precursors to products.

Even systems missing one or the other component can, of course, advance our understanding. Such systems could be called 'infrabiological', because they are not quite biological but are similar to living systems in some crucial respects: elementary combinatorics suggests that out of metabolism (M), boundary (B) and template (T) three dual systems can be built — MT, MB, TB. In particular, coupling of compartment formation with some form of template replication (TB) is the subject of many experiments.

Following earlier work on liposomes (P. Walde, Univ. Zurich), protein expression in these entities has become a viable prospect: liposomes are tiny bags with walls made of layers of phospholipids, like the phospholipids that make up cell membranes. Even composite systems incorporating gene transcription and translation are now possible in liposomes. For example, an artificial stretch of DNA can harbour the gene for T7 RNA polymerase, an enzyme that catalyses the production of RNA from DNA, which in turn induces the expression of green fluorescent protein as an indicator of translation (T. Yomo, Univ. Osaka;



100 YEARS AGO

The Preparation of the Child for Science. A great change in the character of the books concerned with the teaching of science has taken place during the last twenty years or so. A quarter of a century ago the claims of science to a place in the school curriculum were being advocated vigorously, and men of science had still to convince reigning schoolmasters that no education was complete which ignored the growth of natural knowledge and failed to recognise that an acquaintance with the phenomena of nature is necessary to intelligent living. Speaking broadly, it may be said that most classicists even admit now that there are faculties of the human mind which are best developed by practice in observation and experiment. One consequence of the success which has followed the persistent efforts of Huxley and his followers — to secure in the school an adequate recognition of the educative power of science — has been that modern books on science teaching are concerned almost entirely with inquiries into the best methods of instructing young people, by means of practical exercises, how to observe accurately and to reason intelligently. From *Nature* 2 February 1905.

50 YEARS AGO

Principles of Geomorphology. Geomorphology as a science has grown up in the railway age. A hint of what was coming might be espied in those eighteenth-century travellers who, like Gilpin, began very haltingly to display an interest in the form of landscape rather than its formalized versions. A hundred years later and the trains have reached Lucerne; soon we are well into the age of physiography, that pleasant ill-defined compost which made an agreeable part of the later Victorian education. A further hundred years, and this lively branch of science has given birth to a remarkable variety of new and odd words such as pediplains, steptoes and fluviraption... Progress has been rapid; yet the discussion of the characteristics, origin and development of land-forms will long continue to provide an attractive and challenging mental discipline and a valuable education. Geomorphology not only gives scope for the exploratory and cartographical type of mind but also allows abundant opportunity to increase with time the precision of measurement, examination and analysis. Probing, indeed, may gradually replace mapping in this as in other fields. From *Nature* 5 February 1955.

*Towards the Minimal Cell. Erice International School on Complexity, Erice, Sicily, 7–10 December 2004.

K. Tsumoto, Mie Univ., Tsu). The snag is, of course, that these systems contain components taken from contemporary cells, and are far from being self-sufficient.

Replication can also happen in liposomes. RNA from the phage Q β (a virus infecting bacteria) can be incorporated in liposomes (T. Yomo) and be replicated by a replicase enzyme provided by the experimenter. A common by-product of RNA replication is the advent of smaller, faster-replicating mutant RNA molecules, which take over the population. This apparently failed to happen in these experiments, but the reason is debatable. Maybe self-association of template and copy strands reduced competition to such an extent that coexistence is guaranteed (G. von Kiedrowski, Univ. Bochum). Or perhaps the efficient mutants simply failed to arise owing to the small number of replication cycles (E. Szathmáry).

Experimental work is increasingly being complemented by computational investigations. For example, it is possible to account for the growth and fission of compartments in simulations of molecular-assembly dynamics (T. Ikegami, Univ. Tokyo). On the genetic side, the origin of heredity was demonstrated in a simulated system of cross-catalytic autocatalytic networks (K. Kaneko, Univ. Tokyo). Kaneko argued that 'minority control' is a possible origin of heredity in a

bag of genes that constitutes a primordial genome, in that genes with a lower copy number have a more decisive influence on the protocell's simulated behaviour. It is difficult to assess the importance of this finding, as there is no example of the particular network modelled. But the idea may prove helpful in attempts to produce more realistic constructions.

According to the 'composome' model, in which micelles or vesicles are formed from amphiphilic compounds — those having one end that is hydrophilic and the other hydrophobic — there is the prospect of constructing a 'lipid world'. Here, a hereditary component arises from alternative autocatalytic sets of lipids (D. Segré, Harvard Med. School).

Clearly, there is a divide between the top-down and bottom-up approaches, and between theoretical and experimental investigations. In the future, for example, one would like to see more realistic models of the primordial genome and, conversely, an experimental approach to the lipid world. An aim in the coming years will be to bridge those gaps — hence the great value of meetings such as this. ■

*Eörs Szathmáry is at the Collegium Budapest (Institute for Advanced Study), 2 Szenttháromság utca, H-1014 Budapest, Hungary.
e-mail: szathmary@colbud.hu*

reactions and its limitations are known, the situation for reactions at surfaces is much less clear.

In their experiments, White and colleagues¹ prepared nitric oxide molecules in highly excited vibrational states, so that the atoms were subjected to large motion, close to the limit at which the molecules will break up. The excited molecules were scattered from a specially prepared metal surface from which electrons could escape easily. A detector above the surface picked up any electron emission. The experiment's main observation was that when the vibrational energy of the incident nitric oxide molecule exceeded the binding energy of electrons in the surface, electrons were directly emitted from the surface. This finding points to a coupling between nuclear motion and electronic excitation, and therefore indicates that the Born–Oppenheimer approximation is invalid in this case.

The research by White *et al.* extends work in which electronic excitation was produced at metal surfaces by bombardment with various gas-phase species (mostly atoms such as oxygen, hydrogen and nitrogen, high-kinetic-energy rare gases and some molecules)^{2,3}. In one of these experiments³, electrons in the metal tunnelled through a potential-energy barrier to a semiconductor substrate as a result of the bombardment. The charge flow induced in the semiconductor as a result of the tunnelling electrons was termed a 'chemicurrent', to reflect the chemical cause of the electronic excitation.

Although these previous results also point to a breakdown of the Born–Oppenheimer approximation, the situation is somewhat harder to interpret because the electronic excitation is most probably mediated by 'phonons' — vibrational excitations in the substrate itself. White and colleagues' experiment bypasses this poorly defined intermediate step.

Experiments of the type presented by White *et al.* (and the closely related chemicurrent work³) serve as a warning over the widespread use of potential-energy surface models, and should act as an impetus for modifying the conceptual framework used in surface chemistry. There have been attempts to include electronic excitation in theoretical models, but the task is a daunting one and has been limited by a lack of clear experimental findings. The new experiments provide well-characterized results to guide further theoretical development. ■

*Greg Sitz is in the Department of Physics, University of Texas, 1 University Station C1600, Austin, Texas 78712, USA.
e-mail: gositz@physics.utexas.edu*

1. White, J. D., Chen, J., Matsiev, D., Auerbach, D. J. & Wodtke, M. *Nature* **433**, 503–505 (2005).
2. Amirav, A. & Cardillo, M. J. *Phys. Rev. Lett.* **57**, 2299–2302 (1986).
3. Nienhaus, H. *Surf. Sci. Rep.* **45**, 3–78 (2002).

Surface chemistry

Approximate challenges

Greg Sitz

There is growing evidence that the usual approach to modelling chemical events at surfaces is incomplete — an important concern in studies of the many catalytic processes that involve surface reactions.

To describe all the transformations through which a molecule must go during a chemical reaction is a daunting task. The intermediate transition states of a reaction are hard to examine directly, and theory is needed to obtain a full understanding of all the relevant interactions. In 1927, Born and Oppenheimer formulated an 'approximation', which greatly simplified such calculations. Their theory has been crucial to advances in theoretical and chemical physics. It is therefore of great interest when the Born–Oppenheimer approximation breaks down, which may be the case particularly for reactions that take place at surfaces. On page 503 of this issue¹, Jason White and colleagues provide the clearest example to date of such a case.

The break-up of a chemical bond involves a large bond vibration — in other words, a large relative motion of the two atoms that make up the bond. Rather than taking into account all the interactions

involved, the Born–Oppenheimer approximation treats the motion of atomic nuclei separately from electronic excitation. This is justified by the fact that nuclei are much heavier than electrons and move more slowly. Therefore — it is assumed — when nuclei move, as they do during the formation or breaking of a bond, electrons will simply readjust quickly.

Many theoretical methods use this approximation, and solve the Schrödinger equation (the fundamental equation that describes all such interactions) in terms of electrons moving in slowly changing, stationary frameworks of nuclear arrangements. The result can be visualized as a 'potential-energy surface', which plots the solutions of the Schrödinger equation as a function of a molecule's changing structure during a reaction — a popular method for describing chemical reactions. However, although the Born–Oppenheimer approximation has been widely tested for gas-phase

Physiology

A welcome shortage of breath

Thorsten Burmester

The respiratory systems of animals must guarantee an efficient oxygen supply. But it seems that, in some insects, they have evolved to restrict the flow of oxygen too.

Like most other animals, insects need to inhale oxygen and get rid of carbon dioxide. Oxygen fuels energy production in the cells' power plants, the mitochondria, and carbon dioxide is released as a waste product. On page 516 of this issue, Hetz and Bradley¹ show that, contrary to what might be expected, the insect respiratory system may limit rather than assist the uptake of oxygen.

In insects, the exchange of gas with the atmosphere is restricted by a mostly impermeable and inflexible outer layer — the cuticle. Therefore, insects have small openings called spiracles in their cuticle. These are connected to the inner organs by a system of highly branched, gas-filled tubes called tracheae² (Fig. 1, overleaf). Oxygen uptake and carbon dioxide release by the cells mainly occur at the tips of the smallest branches. In highly active organs,

such as flight muscle, the tracheal endings can even enter the cells and reach the mitochondria directly.

A common misconception is that the insect's tracheal system is a very inefficient transport pathway. In fact, oxygen and carbon dioxide are respectively delivered about 200,000 times and 10,000 times faster in tracheal air than in the aqueous environment of the blood^{2,3}. Therefore, simple diffusion through the tracheae would probably be sufficient to supply adequate oxygen and remove carbon dioxide waste even in the largest insects known historically (for example, the dragonfly *Meganeura monyi*, which lived about 280 million years ago and had a wing-span of 70 cm).

The spiracles in the cuticle behave like valves, opening and closing to allow or restrict the insect's gas exchange. Physiologists have long been puzzled by a peculiar

rhythmic respiratory behaviour referred to as the discontinuous gas-exchange cycle (DGC)⁴. In insects exhibiting DGC, the spiracles close for long periods (up to several hours or even days) and open occasionally for only a few minutes. This unusual respiratory pattern has been observed in many adult insects, as well as in resting butterfly and moth pupae. Two main hypotheses have been proposed to explain why some insects display DGC: to reduce water loss through the spiracles, or to adapt to an underground lifestyle. But these ideas were disproved on closer inspection because DGC could be associated with neither the humidity⁵ nor the carbon dioxide concentration⁶ of the environment.

Hetz and Bradley¹ propose a different theory to explain DGC, and their hypothesis has far-reaching implications for how we view animal respiration. They provide compelling evidence that insects use DGC not to acquire but to avoid oxygen. Using the pupae of the moth *Attacus atlas* as a model system, the authors varied the environmental oxygen concentrations from partial pressures of 5 to 50 kPa (the normal atmospheric oxygen partial pressure at sea level is about 21 kPa). Nevertheless, the intra-tracheal oxygen levels in the resting pupae remained low, close to 4 kPa, across the whole range of partial

Atmospheric physics

Seeing the light

An aurora is not just a visual treat for Earth's latitudinally advantaged residents. These events also provide natural laboratories for physicists investigating the complex interplay of electromagnetic waves and ionized particles in plasmas. And we are not merely passive observers of such phenomena — we have the capability of manipulating these processes from the ground, as shown to striking effect by Todd R. Pedersen and Elizabeth A. Gerken elsewhere in this issue (*Nature* **433**, 498–500; 2005).

The playground for these luminous processes is the ionosphere — the ionized upper reach of the atmosphere that stretches from a height of around 100 km to the base of the magnetosphere far above. In Earth's polar regions, the geometry of the geomagnetic field is such that electrons and ions can occasionally be driven down from on high: an aurora (pictured) is the visual manifestation of the collision of these energetic particles with gases in the upper atmosphere and ionosphere.



With this basic understanding of the mechanism in place, researchers have shown previously that it is possible to induce such optical processes artificially. By pumping high-power radio waves into the ionosphere (the frequency of the waves being tuned to the local plasma environment), electrons can be locally energized to collide with atmospheric gases in a manner analogous to the natural auroral process. But the resulting optical effects are small, with emission intensities falling well below the detection limit of the human eye.

Now Pedersen and Gerken have shown that, if ionospheric conditions are just right, much stronger emissions can be generated by this approach — so strong, in fact, that they are in principle visible to the naked eye. The 'trick' underlying this demonstration was to choose a time when a natural aurora was already active (previous efforts were directed at quiet regions of the ionosphere). And rather than targeting the main ionospheric layer, the researchers tuned the radio waves to excite a much lower layer, which had been transiently ionized

by the inbound charged particles.

These observations raise many questions about the processes involved. For example, is an active aurora really a prerequisite for generating such bright emissions, or is it instead largely a picturesque bystander? Should the answer turn out to be the latter, we are left with the tantalizing (some would say disconcerting) possibility that such radio-fuelled emissions could form the basis of a technology for urban lighting, celestial advertising, and more...

Karl Ziemelis

E. KENNEDY/NAVAL RESEARCH LAB.

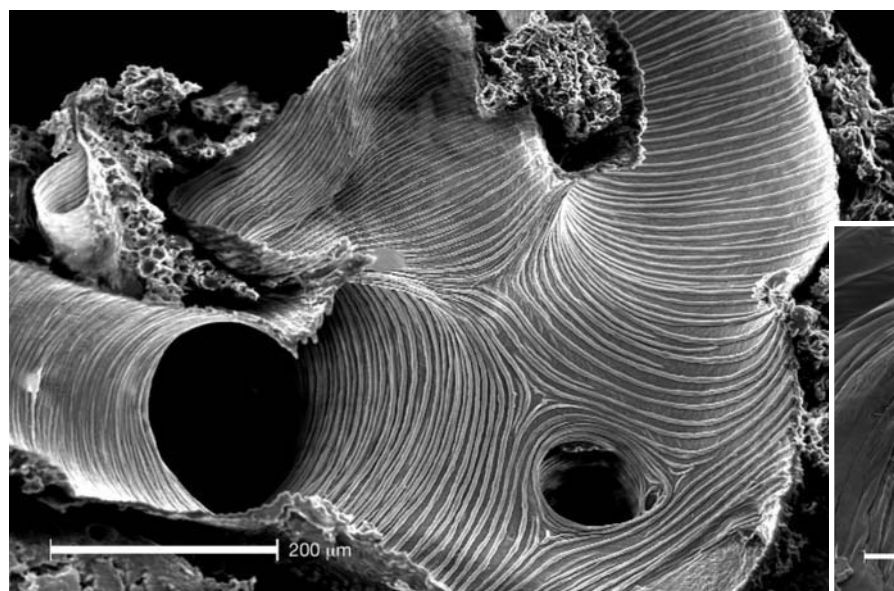
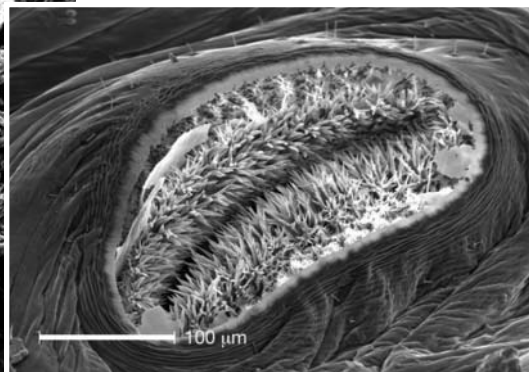


Figure 1 Insect breathing apparatus. Left, the tracheal system of the beetle *Zophobas rugipes*; below, a spiracle of the same species.



pressures. Thus, the moth pupa limits the amount of oxygen taken in by keeping the spiracles closed for as long as possible, and opening them only to get rid of the accumulated carbon dioxide.

At first glance, the idea that an air-breathing animal should try to limit apparently normal oxygen levels seems perplexing. But oxygen is a double-edged sword: although required to fuel energy production, it is also a potent source of toxic compounds known as reactive oxygen species (ROS), which can damage proteins, DNA and lipids⁷. In recent years, ROS have been recognized as a major threat to cell survival, and toxic ROS effects are suggested to underlie ageing and cell death. Therefore, it is advantageous to keep cellular oxygen levels just high enough for efficient mitochondrial respiration, and as low as possible to minimize oxidative damage. Obviously, the critical oxygen concentration in moth pupae is far below the normal atmospheric level of about 21%. This is probably true for other animals too; for instance, quite low oxygen levels (0.4–5 kPa) are also found in mammalian tissues⁸.

But if atmospheric oxygen concentrations are toxic to resting pupae, why aren't they noxious to the insects that rarely or never close their spiracles? The answer probably lies in differences in metabolic activity. The insects' tracheal system is well designed for efficient oxygen supply during periods of high activity, when oxygen never accumulates to critical concentrations in the cell because it is rapidly converted into water by the respiratory chain. However, in periods when respiration falls, such as in the resting butterfly pupa, oxygen consumption is too low to prevent it building up to harmful levels. It seems that a particular breathing pattern, the DGC, has evolved to ensure oxygen homeostasis. ■

Thorsten Burmester is at the Institute of Zoology,

University of Mainz, Müllerweg 6, D-55128 Mainz, Germany.

e-mail: burmeste@uni-mainz.de

1. Hetz, S. K. & Bradley, T. J. *Nature* **433**, 516–519 (2005).
2. Kestler, P. *Environmental Physiology and Biochemistry of Insects* (ed. Hoffmann, K. H.) 137–186 (Springer, Berlin, 1985).
3. Krogh, A. *Pflügers Arch. Ges. Physiol.* **179**, 95–120 (1920).
4. Lighton, J. R. B. *Annu. Rev. Entomol.* **41**, 309–324 (1996).
5. Quinlan, M. C. & Hadley, N. F. *Physiol. Zool.* **66**, 628–642 (1993).
6. Gibbs, A. G. & Johnson, R. A. *J. Exp. Biol.* **207**, 3477–3482 (2004).
7. Halliwell, B. & Gutteridge, J. M. C. *Free Radicals in Biology and Medicine* (Oxford Univ. Press, 1999).
8. Vanderkooi, J. M., Erecinska, M. & Silver, I. A. *Am. J. Physiol. Cell Physiol.* **260**, C1131–C1150 (1991).

RNA interference

Methylation mystery

Michael Ronemus and Rob Martienssen

Tiny RNA molecules called microRNAs are important in development, and are thought to function by causing the degradation of matching messenger RNAs. That may not be their only mode of action, however.

RNA molecules come in various sizes, ranging from the very long to the very short. The smaller RNAs fall into two major classes: microRNAs (miRNAs), which guide development in an organism by regulating target genes¹; and small interfering RNAs (siRNAs), which target viruses, inserted genes and mobile genetic elements — a significant function being defence of the genome². One way in which siRNAs work is by guiding the modification (by methylation) of DNA strands, as well as the modification of the histone proteins around which DNA is wrapped, thus silencing gene expression³. Writing in *Developmental Cell*, Bao *et al.*⁴ suggest that miRNAs might also — contrary to expectation — contribute to DNA methylation.

At a casual glance, miRNAs and siRNAs don't seem all that different. Both are short, single-stranded RNA molecules, generally 21–22 nucleotides in size. Both are processed from true or transiently double-stranded precursor RNA molecules, by specialized enzymes called Dicers. And both programme the activity of Argonaute proteins in RNA-

induced silencing complexes (RISCs)⁵.

But look more closely, and it seems that any similarity is nothing more than a consequence of an ancient common origin. An miRNA precursor originates from a non-protein-coding gene and is processed in a multi-step pathway that is coupled to its export from the cell nucleus. Once channelled to the cytoplasm, the miRNA interacts with a target messenger RNA (mRNA), via the RISC, to trigger the destruction of the mRNA, or to prevent it from being translated into protein¹. This, then, effectively silences the gene that encodes the mRNA.

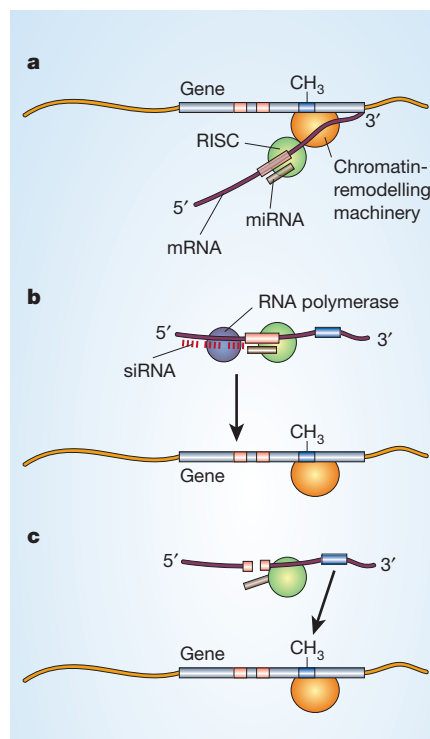
In contrast, the precursor of an siRNA is any double-stranded RNA molecule, a feature that allows for multiplication of the original siRNA through reiterative dicing and synthesis⁶. siRNAs target complementary RNAs for degradation and also influence the modification of repetitive DNA and the proteins that package it. Adding to the differences between the two RNAs, certain miRNA sequences are conserved throughout the animal kingdom or between distantly related plant species, whereas siRNA

sequences vary widely⁷. Also, miRNAs are almost never exact matches of their targets, whereas siRNAs are perfectly complementary to theirs⁵.

It now seems, however, that the functional distinction between miRNAs and siRNAs is rather more blurred than was thought. Using *Arabidopsis* plants, Bao *et al.*⁴ have examined mutated variants of *PHABULOSA* (*PHB*) and *PHAVOLUTA* (*PHV*) — genes whose expression is usually reduced by miRNAs. These genes were known to act redundantly to specify upper (adaxial) and lower (abaxial) polarity in lateral organs such as leaves. The mutations in question result in an even less perfect match between the *PHV* or *PHB* mRNAs and their cognate miRNAs¹. So, for instance, the miRNA that targets *PHB* is usually expressed only on the abaxial sides of *Arabidopsis* leaves⁸, preventing *PHB* expression there and allowing these regions to adopt an abaxial fate. But when *PHB* is mutated, the levels of its mRNA increase — as would be expected if, because of the imperfect matching, the mRNA is no longer being destroyed — and *PHB* expression expands into the abaxial domain. Leaves then adopt inside-out (adaxialized) fates⁹.

At least, that's how the mutations were thought to work. But Bao and colleagues' observations may complicate this interpretation. In wild-type plants, the authors found extensive DNA methylation over a region of several hundred base pairs, located more than 1 kilobase downstream of the miRNA-binding sites in the *PHB* and *PHV* genes. The methylation was severely reduced when the miRNA-binding sites in these genes were mutated, leading Bao *et al.* to suggest that this reduction in methylation helps to 'desilence' the genes, and that such methylation is therefore dependent on miRNAs. The demethylation shows various hallmarks of RNA-mediated silencing in plants¹⁰. And other observations (for instance, the miRNA spans a 'splice junction' in the cognate mRNA) suggest that a direct interaction between the DNA and miRNA is unlikely, and that binding of the miRNA to mRNA is required for methylation to take place⁴.

Bao *et al.* suggest a simple explanation for how the proposed miRNA-induced DNA methylation might occur. The miRNA, presumably located in the cytoplasm, re-enters the nucleus and pairs with the mRNA while the latter is still closely associated with its encoding DNA (Fig. 1). The DNA then becomes methylated, downstream of the *PHB* or *PHV* sequences that match the miRNA, by an unspecified 'chromatin-remodelling complex', which is guided either by the miRNA itself or by cleavage of the mRNA. Obvious candidates for such a complex are a variant of the RISC that incorporates the ARGONAUTE1 protein^{1,8,9,11}, or a complex that interacts with it.



But Bao *et al.* found no effect of either ARGONAUTE1 or its close relative PINHEAD on methylation at the *PHB* locus. And a more general problem with this model is that plants lacking one particular miRNA-processing Dicer enzyme also show no change in methylation of *PHB*⁴, so severe reductions in the levels of processed miRNAs do not limit DNA methylation. It also remains unclear how a miRNA would target methylation specifically — and only — to a small region 1–2 kilobases from any cognate sequence.

An alternative model is that the binding of miRNA to its target mRNA sets in motion a secondary cascade of RNA-mediated silencing, culminating in the methylation of DNA sequences homologous to the mRNA. The likely mediators would be secondary siRNAs whose production is initiated by the miRNA, but which emanate from the copying of other portions of the target mRNA (Fig. 1). There is in fact a precedent for this¹², and an intermediate step involving siRNAs could account both for the presence of DNA methylation and, via methylation 'spreading', for its location far from the miRNA-target site^{3,13}. However, methylation is not found between this location and the miRNA-target site, arguing against spreading. Another possibility is that the miRNA somehow modifies the mRNA following its cleavage, allowing the degradation products to target the DNA for methylation. But there are no data to support either model.

So the mechanism remains a mystery. Yet it is also possible that DNA methylation is only a minor (or redundant) player in the regulation of *PHB* and *PHV*. Indeed, when the cellular pathways that produce this type of DNA methylation are mutated, the effects

on the plant are not the same as those produced by the *PHB* mutation studied^{4,14}, and methylation of certain other plant genes does not seem to affect their expression⁷. Rather than silencing the gene, the methylation might serve as an imprint, established by the miRNA signal but 'remembered' later in development, or even in subsequent generations (although other evidence argues against such stability^{4,8}, and against the need for such an imprint⁸). But it may be hard to distinguish between cause and effect: in plants, variable methylation patterns — even those induced by mutagenesis — can be maintained over many generations¹⁵, perhaps explaining the loss of methylation in the *PHB* mutants. Whatever the mechanism, the parallels between miRNA-dependent DNA methylation and siRNA-induced gene silencing are certainly intriguing enough to warrant further investigation.

Michael Ronemus and Rob Martienssen are at the Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.

e-mail: martienssen@cshl.edu

1. Bartel, D. P. *Cell* **116**, 281–297 (2004).
2. Lippman, Z. & Martienssen, R. *Nature* **431**, 364–370 (2004).
3. Matzke, M. *et al.* *Biochim. Biophys. Acta* **1677**, 129–141 (2004).
4. Bao, N., Lye, K.-W. & Barton, M. K. *Dev. Cell* **7**, 653–662 (2004).
5. He, L. & Hannon, G. J. *Nature Rev. Genet.* **5**, 522–531 (2004).
6. Baulcombe, D. *Nature* **431**, 356–363 (2004).
7. Lippman, Z. *et al.* *Nature* **430**, 471–476 (2004).
8. Kidner, C. A. & Martienssen, R. A. *Nature* **428**, 81–84 (2004).
9. McConnell, J. R. & Barton, M. K. *Development* **125**, 2935–2942 (1998).
10. Chan, S. W. *et al.* *Science* **303**, 1336 (2004).
11. Fagard, M. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 11650–11654 (2000).
12. Parizotto, E. A., Dunoyer, P., Rahm, N., Himber, C. & Voinnet, O. *Genes Dev.* **18**, 2237–2242 (2004).
13. Jones, L. *et al.* *Curr. Biol.* **11**, 747–757 (2001).
14. Cao, X. & Jacobsen, S. E. *Proc. Natl Acad. Sci. USA* **99** (Suppl. 4), 16491–16498 (2002).
15. Stokes, T. L., Kunkel, B. N. & Richards, E. J. *Genes Dev.* **16**, 171–182 (2002).

Physics

Cool gas, hotter superconductors

Science doi:10.1126/science.1109220 (2005)

J. Kinast *et al.* have obtained strong evidence that an extremely cold gas of lithium-6 atoms can become a unique type of superfluid.

Lithium-6 atoms are fermions, which cannot share the same quantum state and hence do not readily become a superfluid, a medium that flows without friction. But they can form strongly bound non-molecular pairs — analogues of the Cooper pairs of electrons responsible for superconductivity, where current flows without resistance.

Previously, the researchers had seen this 'Fermi gas' of tightly bound atom pairs vibrating in a way that suggested it could become a superfluid at sufficiently low temperatures. Now, they have studied the thermodynamics of a Fermi gas of lithium-6 atoms. They noted a clear jump in the heat capacity after precisely reducing the energy of the gas, consistent with a phase change signalling the creation of a superfluid. The findings also coincide with the predictions of a theory the authors initially created for high-temperature superconductors.

Kinast *et al.* say that understanding this superfluid state could help to develop superconductors that operate well above room temperature.

Mark Peplow

Evolutionary biology

Reinvention of a nose

Curr. Biol. **15**, 116–121 (2005)

Despite being descended from marine ancestors, the robber crab *Birgus latro* would drown in water. But it has no fear of heights, and can easily climb to the tops of palm trees to break open coconuts with its massive claws. Life on land has required other adaptations as well, and Marcus C. Stensmyr *et al.* have cracked the question of how the olfactory system of these giant crabs has become so well suited to terrestrial living.

Stensmyr *et al.* first verified the reputed ability of robber crabs to track down any smelly food in their immediate environment. The authors then looked at the crab's olfactory adaptations. They found that the robber crab has sensitive carbon dioxide detectors, which, together with other odour sensors, help it to locate food sources. Moreover, the bristle hairs on the animals' antennules have characteristics that enable odours to be sensed while minimizing water evaporation.

Strikingly, recordings of the output from the crab's olfactory neurons show strong similarities to those from insects. This, the



Phylogeny

Hippo relations

Proc. Natl Acad. Sci. USA **102**, 1537–1541 (2005)

A morphological analysis has helped to fill a large gap in the evolutionary story of the hippopotamus. The discovery also brings researchers a step nearer to closing the book on a debate that has lasted more than 150 years.

Taxonomists had suggested that the nearest living relatives of hippos are pigs. But genetic analyses indicated that hippos are more closely related to cetaceans (whales, dolphins and the like).

Jean-Renaud Boisserie *et al.* now report that hippos are the only surviving members of a group of animals known as anthracotheres — and that the anthracotheres are the sister group of the cetaceans. They base this conclusion on studies of a range of animal species, including *Libycosaurus*, a semi-aquatic anthracothere that lived in Africa between 12 million and 6 million years ago.

The link between hippos and anthracotheres, which are well represented in the fossil record, also enables the evolutionary history of hippos to be traced back through what was a frustrating 40-million-year hole in their story. **Michael Hopkin**

researchers explain, makes the crab and insect smelling systems a striking example of convergent adaptation.

Roxanne Khamisi

Developmental biology

Four-square specialist

Development **132**, 479–490 (2005)

Biologists have thought that each cell in a four-cell mammalian embryo has exactly the same ability to give rise to tissues as its neighbour. Karolina Piotrowska-Nitsche *et al.* challenge this idea, showing that each cell tends to behave differently.

The authors have found before that, in most two-cell mouse embryos, one of the cells divides along a vertical line that runs from top to bottom (meridionally). The other divides roughly along a horizontal line that spans its equator (equatorially).

Piotrowska-Nitsche *et al.* have now made 'chimaeric' embryos by sandwiching together cells of the same type. When all the cells were derived from equatorial divisions, the embryos suffered developmental problems and some did not survive to birth. When all the cells were derived from meridional divisions, the embryos developed normally.

The researchers went on to show that a single cell derived from an equatorial division, when added to four randomly selected cells, does have the ability to give rise to all tissue types. But the results suggest that, during normal embryo development, cells start specializing from an earlier stage than was thought.

Helen Pearson

Atmospheric physics

TIGER burning bright

Geophys. Res. Lett. **32**, L02801 (2005)

During thunderstorms, the air is full of sprites and elves. These are the evocative names of flashes of light in the upper atmosphere, known as 'transient luminous events' (TLEs). The cause is excitation of gas molecules following a lightning discharge: elves are triggered by the electromagnetic pulse from cloud-to-ground lightning; sprites and other types of TLE may follow within a fraction of a second, typically as red or blue flashes arising well above the thundercloud.

In January 2003, the crew of the space shuttle Columbia recorded a decidedly odd TLE, as Yoav Yair *et al.* now report. The observation was made by astronaut Ilan Ramon — one of the crew killed in the shuttle's ill-fated final flight just days later — as Columbia passed just south of Madagascar.

A lightning discharge occurred near the Earth's limb — the visible edge of its atmosphere. It was immediately followed by a TLE to the east, more than 1,000 km from the thunderstorm. The shape of this flash was unlike that of known TLEs, and its separation in time and space from the lightning also made it unusual. The researchers call it a 'transient ionospheric glow emission in red' (TIGER), and it seems that these ghostly apparitions may not require a nearby thunderstorm.

Philip Ball

Burrow extension by crack propagation

A worm minimizes its energy expenditure as it forges a path through mud sediment.

Until now, the analysis of burrowing mechanics has neglected the mechanical properties of impeding, muddy, cohesive sediments, which behave like elastic solids¹. Here we show that burrowers can progress through such sediments by using a mechanically efficient, previously unsuspected mechanism — crack propagation^{1,2} — in which an alternating ‘anchor’ system of burrowing serves as a wedge to extend the crack-shaped burrow. The force required to propagate cracks through sediment^{1,2} in this way is relatively small: we find that the force exerted by the annelid worm *Nereis virens* in making and moving into such a burrow amounts to less than one-tenth of the force it needs to use against rigid aquarium walls³.

Using gelatin as an analogue for sediment (on the basis of its similar mechanical properties)^{1,2,4} and polarized light to visualize the burrow around *N. virens*, we investigated whether these worms burrow by crack propagation. Gelatin is birefringent, which enables its stress to be analysed⁵ during burrowing.

We found that the edges of the burrow were visible and that they showed evidence of a discoidal crack, which was held open by the dorsal and ventral surfaces of the animal (Fig. 1a,b). The arc of the crack extends laterally beyond the setae and anteriorly to the tips of the antennae. Visualization of stress fields by photoelastic stress analysis (for methods, see supplementary information) reveals a force exerted dorsoventrally by eversion of the animal’s pharynx against the relatively flat wall of the crack (Fig. 1c).

According to linear elastic fracture mechanics, the stress intensifies at the crack tip, propagating it when the critical stress intensity is reached^{1,6}. Here, the everted pharynx acts as a wedge, with the radial force intensified at the tip of the crack, which propagates, allowing the animal to move forward (Fig. 2). The crack’s aspect ratio (T/W in Fig. 2) is a function of the material properties of the sediment (the critical stress intensity and Young’s modulus). Expansion of a subterminal body portion has previously been considered as an anchor⁷; we infer instead that a terminal or subterminal lateral expansion acts as a wedge in elastic mud.

We measured the maximum force required to propagate a crack as 0.023 ± 0.002 newtons (mean $\pm$ s.e.m.; $n = 5$), which is less than 10% of the ‘radial’ force of 0.6 N that is exerted by a worm against an aquarium wall³. Sediment behaves like an elastic solid that deforms under stress, as evidenced by acoustic⁸ and bubble-growth experiments¹. The relatively small forces exerted by growing bubbles in

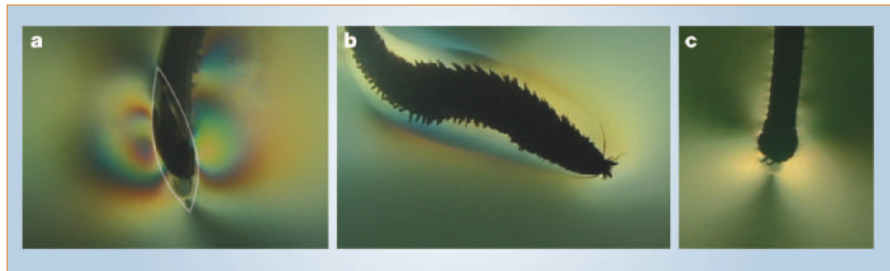


Figure 1 The crack-shaped burrow around the annelid worm *Nereis virens* in polarized light. **a**, Anterior view in crosspolarized light, showing the longer axis of the discoidal crack, which is parallel to the setae, and showing the dorsal and ventral surfaces against the gelatin substrate; the shape of the burrow follows the line encircling the crack. **b**, Dorsal view in crosspolarized light, showing the shape of the burrow around the animal. **c**, Side view in circularly polarized light, showing the stress field dorsal and ventral to the worm, and the crack extending anteriorly.

sediment are well predicted by linear elastic fracture mechanics from the mechanical properties of the medium¹. According to the theory of elasticity, a closer, rigid boundary necessitates greater force for a given displacement, resulting in substantial effects of nearby rigid walls on burrowers.

Crack propagation may also be a likely mechanism of movement for burrowers from several other marine phyla. For example, clams and echinoids, which are both found in muddy sediments, could exploit their wedge-like shape — and some echinoids have been reported to push directly into the frontal sediment, rather than excavating it, and to move through the sediment by means of a repeated rocking motion, unlike more globular, excavating echinoids living in sands⁹. Gammarid amphipods could exploit their resemblance to oblate bubbles^{1,2}, and subterminal expansions in earthworms indicate that they may burrow, and roots grow¹⁰, by an analogous mechanism.

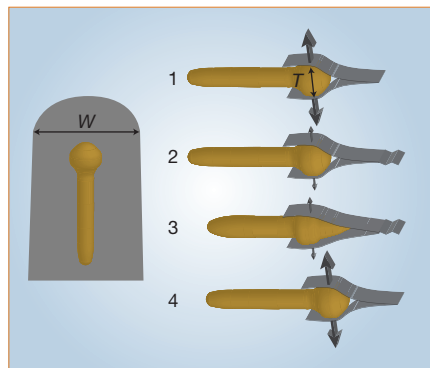


Figure 2 Scheme showing dorsal view of crack shape and lateral oblique views of crack propagation during burrowing. Arrows extending vertically from the crack indicate forces. The worm everts its pharynx and exerts a force normal to the direction of movement (1), which causes the crack to propagate, releasing energy (2). The worm then retracts the pharynx and moves forward into the crack (3) before repeating the cycle (4). W is the longer axis of the discoidal crack; T , the shorter.

Burrowing by crack propagation also bears on bioturbation, the movement of sediment grains and fluids by living organisms. Bioturbation models have generally ignored the polymeric matrix that leads to elastic behaviour and that keeps sediment grains in the same relative position to each other, instead considering particles as separate, ‘randomly wandering’ elements¹¹. But cracks open new surfaces from which animals can feed, leading to unsteady release and uptake of solutes. Crack propagation focuses stresses at the propagating crack tip. These stresses are probably the highest experienced by sediments after their deposition and bear on issues such as the permanence of clay–grain associations and the mechanical protection of organic material from decay.

Kelly M. Dorgan*, **Peter A. Jumars***,
Bruce Johnson†, **B. P. Boudreau†**,
Eric Landis‡

*Darling Marine Center, University of Maine,
Walpole, Maine 04573, USA

e-mail: kelly.dorgan@umit.maine.edu

†Department of Oceanography, Dalhousie
University, Halifax, Nova Scotia B3H 4J1, Canada

‡Department of Civil and Environmental
Engineering, University of Maine, Orono,
Maine 04469, USA

1. Johnson, B. D., Boudreau, B. P., Gardiner, B. S. & Maass, R. *Mar. Geol.* **187**, 347–363 (2002).
2. Boudreau, B. P. *et al. Geology* (in the press).
3. Hunter, R. D. & Elder, H. Y. *J. Zool.* **218**, 209–222 (1989).
4. Menand, T. & Talt, S. R. *Nature* **411**, 678–680 (2001).
5. Full, R. J., Yamauchi, A. & Jindrich, D. L. *J. Exp. Biol.* **198**, 2441–2452 (1995).
6. Broek, D. *Elementary Engineering Fracture Mechanics* (Sijthoff & Noordhoff, Alphen aan den Rijn, the Netherlands, 1978).
7. Elder, H. Y. in *Aspects of Animal Movement* (ed. Trueman, E. R.) 71–92 (Cambridge Univ. Press, Cambridge, 1980).
8. Hamilton, E. L. *J. Acoust. Soc. Am.* **68**, 1313–1340 (1980).
9. Kanazawa, K. *Palaeontology* **35**, 733–750 (1992).
10. Abdalla, A. M., Hettiaratchi, D. R. P. & Reece, A. R. *J. Agric. Eng. Res.* **14**, 236–248 (1969).
11. Meysman, F. J. R., Boudreau, B. P. & Middelburg, J. J. *J. Mar. Res.* **61**, 391–410 (2003).

Supplementary information accompanies this communication on Nature’s website.

Competing financial interests: declared none.

Nanotechnology

'Buckypaper' from coaxial nanotubes

Double-walled carbon nanotubes are needed in a pure, highly crystalline form before features such as their electronic properties, thermal transport and mechanical behaviour can be investigated. Here we fabricate a paper-like material that consists of hexagonally packed bundles of clean, coaxial carbon nanotubes whose double walls vary little in diameter; it is prepared in high yields using chemical-vapour deposition with a conditioning catalyst and two-step purification. Our results will enable investigation of the physical properties of double-walled carbon nanotubes, which are predicted to be superior to those of both their single- and multi-walled relatives.

Double-walled carbon nanotubes, or DWNTs, consist of two concentric graphene cylinders, a structure that is intermediate between single-walled and multiwalled carbon nanotubes (SWNTs and MWNTs, respectively). Because they may have striking new electronic and mechanical properties, several attempts have been made to produce very pure DWNTs^{1–7}; however, these have typically generated a mixture of DWNTs and SWNTs that is contaminated with metal particles, amorphous carbon and multilayer carbon nanotubes. Catalytic chemical-vapour deposition is generally considered to be the most efficient method for the large-scale production of nanotubes^{8,9}; arc-discharge methods and thermal treatments of peapods of C₆₀ (buckminsterfullerene, or 'buckyballs') result in unwanted carbon nanomaterials as well as DWNTs^{1,4}. But measuring the cross-section of such DWNTs by high-resolution transmission electron microscopy shows that their uniformity and purity fail to approach those of SWNTs^{1–7}.

We synthesized DWNTs using a conditioning catalyst of molybdenum at one end of the furnace and a nanotube catalyst of iron in the central part of the furnace (for details of methods, see supplementary information). A methane–argon gas mixture (1:1) was then fed into the reactor for 10 min at 875 °C. Use of the conditioning catalyst¹⁰ caused more DWNTs than SWNTs to grow, possibly because of an increase in the amount of active carbon species.

To obtain sheets of pure DWNT paper, we applied a two-step purification process to the synthesized products. Hydrochloric acid (18% by weight HCl at 100 °C for 10 h) was used to remove the iron catalyst and supporting material, followed by oxidation in air at 500 °C for 30 min, which removed amorphous carbon and chemically active SWNTs. After filtration, we obtained a dark, paper-like sheet, which was very flexible

and mechanically stable (Fig. 1a).

Transmission electron microscopy revealed that the yield of DWNTs was extremely high (more than 95%) and that they were arranged in bundles (Fig. 1b–d). The DWNT paper consists of nanotubes with two remarkably consistent diameters that are hexagonally packed (Fig. 1d). Raman spectra verified the tube-diameter distributions from the radial breathing-mode (RBM) frequencies (see supplementary information),

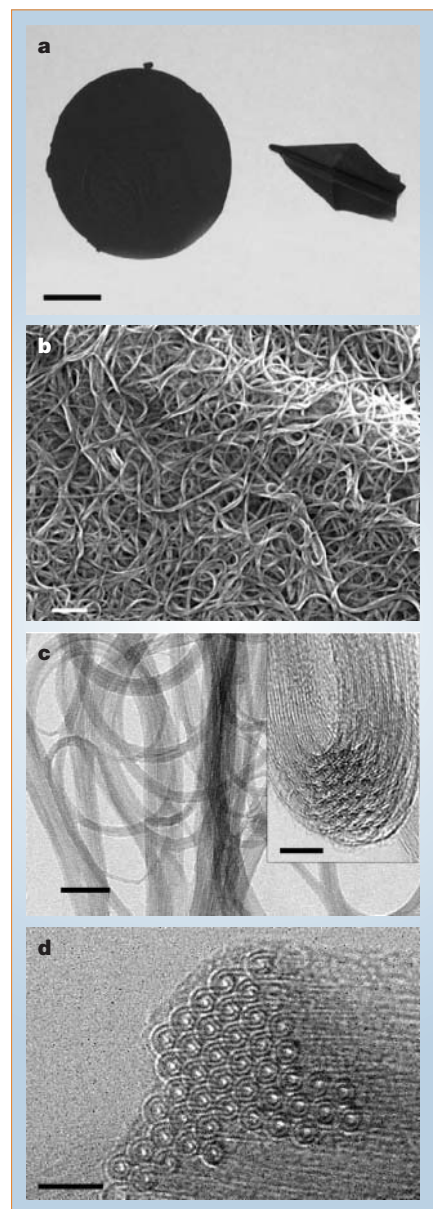


Figure 1 Production of pure and clean double-walled carbon nanotubes (DWNTs) in high yield. **a**, Photograph of double-walled carbon nanotube 'buckypaper'; the paper (left) is tough and flexible enough to fold into an origami plane (right). **b**, Scanning electron micrograph and **c**, low-magnification transmission electron micrograph (TEM; JEOL JEM-2010FEF) of DWNT paper, showing bundles of carbon nanotubes (insert, a single bundle; note the material's resistance to bending). Impurities are notably absent in these images. **d**, Typical high-resolution TEM image of a different bundle, showing the perfect hexagonal packing structure in both cross-section and side view. Scale bars: **a**, 1 cm; **b**, 300 nm; **c** 50 nm; inset, 5 nm; **d**, 5 nm.

which are inversely related to tube diameter. Raman peaks appear above 250 cm^{−1} (corresponding to the inner shells of DWNTs) and below 250 cm^{−1} (usually associated with the outer shells of DWNTs). Using the equation $\omega_{\text{RBM}} = 234/d_t + 10$, where d_t is the tube diameter (in nanometres) and ω_{RBM} is the RBM frequency (in cm^{−1})¹¹, we were able to define two sets of DWNT pairs having inner-to-outer diameter ratios of 0.77:1.43 and 0.90:1.60, respectively.

Now that this pure material is available, it will be possible to investigate whether DWNTs behave as quantum wires and whether there is a chirality relationship between concentric tubes during growth, as well as the effect of tube concentricity on electronic conductance and the adsorption properties of coaxial nanotube ropes. This material should be useful in the fabrication of, for example, nanocomposites, field emission sources, nanotube bi-cables and electronic devices^{12,13}. It is possible that DWNTs will eventually replace SWNTs or MWNTs in various applications because their mechanical properties, thermal conductivity and structural stability are likely to be superior owing to their coaxial structure.

M. Endo*, H. Muramatsu*, T. Hayashi*, Y. A. Kim*, M. Terrones†, M. S. Dresselhaus‡

*Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano-shi 380-8553, Japan e-mail: endo@endomoribu.shinshu-u.ac.jp

†Advanced Materials Department, IPICYT, Camino a la Presa San José 2055,

San Luis Potosí 78216, Mexico

‡Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA

- Bandow, S., Takizawa, M., Hirahara, K., Yudasaka, M. & Iijima, S. *Chem. Phys. Lett.* **337**, 48–54 (2001).
- Wei, J. et al. *J. Mater. Chem.* **13**, 1340–1344 (2003).
- Lyu, S. C. et al. *Chem. Mater.* **15**, 3951–3954 (2003).
- Sugai, T. et al. *Nano Lett.* **3**, 769–773 (2003).
- Hiraoka, T. et al. *Chem. Phys. Lett.* **382**, 679–685 (2003).
- Flahaut, E., Bacsar, R., Peigney, A. & Laurent, C. *Chem. Commun.* 1442–1443 (2003).
- Lyu, S. C. et al. *J. Phys. Chem. B* **108**, 2192–2194 (2004).
- Oberlin, A., Endo, M. & Koyama, T. *J. Cryst. Growth* **32**, 335–349 (1976).
- Endo, M. *Chem. Tech.* 568–576 (1988).
- Franklin, N. R. & Dai, H. *Adv. Mater.* **12**, 890–894 (2000).
- Jorio, A. et al. *Phys. Rev. Lett.* **86**, 1118–1121 (2001).
- Kurachi, H. et al. *Proc. 21st Intl Display Res. Conf./8th Intl Display Workshops* 1237–1240 (2001).
- Endo, M. et al. *Nano Lett.* **4**, 1451–1454 (2004).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

brieF communications arising online

www.nature.com/bca

Biochemistry: Role of PQQ as a mammalian enzyme cofactor?

L. M. Felton, C. Anthony (doi:10.1038/nature03322)

Biochemistry: Is pyrroloquinoline quinone a vitamin?

R. Rucker, D. Storms, A. Sheets, E. Tchapanian, A. Fascetti (doi:10.1038/nature03323)

Biochemistry: Reply

T. Kasahara, T. Kato (doi:10.1038/nature03324)

Biochemistry

Role of PQQ as a mammalian enzyme cofactor?

Arising from: T. Kasahara & T. Kato *Nature* **422**, 832 (2003)

The announcement by Kasahara and Kato of a new redox-cofactor vitamin for mammals¹, pyrroloquinoline quinone (PQQ), was based on their claim that an enzyme, predicted to be involved in mouse lysine metabolism, is a PQQ-dependent dehydrogenase. However, this claim was dependent on a sequence analysis using databases that inappropriately label β -propeller sequences as PQQ-binding motifs. What the evidence actually suggests is that the enzyme is an interesting novel protein that has a seven-bladed β -propeller structure, but there is nothing to indicate that it is a PQQ-dependent dehydrogenase.

In bacteria, PQQ is an essential cofactor for various dehydrogenase enzymes known as quinoproteins². Nutritional experiments have indicated some (unknown) metabolic or nutritional role for PQQ in mammals^{1,3,4}, but it cannot be accepted as a vitamin until it is proved to be required by an enzyme as an essential cofactor; this is the key evidence that Kasahara and Kato¹ claim to have provided.

In the course of a study on bipolar disorder (see www.brain.riken.go.jp/labs/mdmd/pqq/index-e.html), these authors cloned a mouse gene encoding a protein (U26) with some similarity to yeast aminoadipate reductase (LYS2)⁵; they proposed that mouse U26 could be involved in one of the important first steps in the degradation of dietary lysine, acting as a PQQ-dependent adipic 6-semialdehyde dehydrogenase.

As would be expected from the method used by Kasahara and Kato in searching for the LYS2 analogue¹, the U26 sequence contained no carboxy-terminal NAD(P)-binding domain. They noted from sequence analysis that U26 has an alternative carboxy-terminal domain that contains seven repeats of the 'PQQ-binding motif' that is conserved among bacterial PQQ-dependent dehydrogenase enzymes, leading to the conclusion that mouse U26 could be a PQQ-dependent dehydrogenase.

However, this conclusion is based on the misconception that the Smart and Pfam

databases are able to recognize PQQ-binding sites even when, as in this case, there is negligible sequence similarity to known PQQ-dependent enzymes. The 'sites' wrongly identified by the databases do not represent PQQ-binding sites but represent the β -sheets that form the 'blades' of the 'propeller fold' that happens to be a feature of all PQQ-dependent dehydrogenases, whose main structure is a superbarrel made up of either six or eight 'propeller blades' (Fig. 1). The propeller fold is not related in any direct way to PQQ binding^{2,6}, and these folds are found in many other types of protein, which have extreme functional and phylogenetic diversity⁷.

We contend that there is still no compelling evidence for a PQQ-dependent enzyme in the mouse and that the authors' announcement of a new vitamin was therefore premature.

Leigh M. Felton, Chris Anthony

School of Biological Sciences, University of Southampton, Southampton SO16 7PX, UK
e-mail: c.anthony@soton.ac.uk

doi:10.1038/nature03322

1. Kasahara, T. & Kato, T. *Nature* **422**, 832 (2003).
2. Goodwin, P. M. & Anthony, C. *Adv. Microb. Physiol.* **40**, 1–80 (1998).
3. McIntire, W. S. *Annu. Rev. Nutr.* **18**, 145–177 (1998).
4. Steinberg, F. et al. *Exp. Biol. Med.* **228**, 160–166 (2003).
5. Ehmann, D. E., Gehring, A. M. & Walsh, C. T. *Biochemistry* **38**, 6171–6177 (1999).
6. Anthony, C. & Ghosh, M. *Progr. Biophys. Mol. Biol.* **69**, 1–21 (1998).
7. Paoli, M. *Progr. Biophys. Mol. Biol.* **76**, 103–130 (2001).

Reply: T. Kasahara and T. Kato reply to this communication (doi:10.1038/nature03324).

Biochemistry

Is pyrroloquinoline quinone a vitamin?

The announcement by Kasahara and Kato¹ of pyrroloquinoline quinone (PQQ) as a 'new' vitamin has received considerable attention. We have since attempted to reproduce the findings on which their conclusion is based, namely that defects in lysine metabolism occur in PQQ-deprived rodents. However, we find that the activity of α -aminoadipic acid- δ -semialdehyde (AAS) dehydrogenase in liver and plasma levels of α -aminoadipic acid (AAA), both of which act as indicators of lysine degradation in mammals, are not affected by changes in PQQ dietary status. Our results call into question the identification of PQQ as a new vitamin.

The observations of Kasahara and Kato¹ were of particular interest to us because their experimental model was based on our findings that mice that are fed chemically defined diets and deprived of PQQ often show signs of reproductive failure and compromised neonatal growth^{2,3}. Using previously described assay conditions⁴, we found that

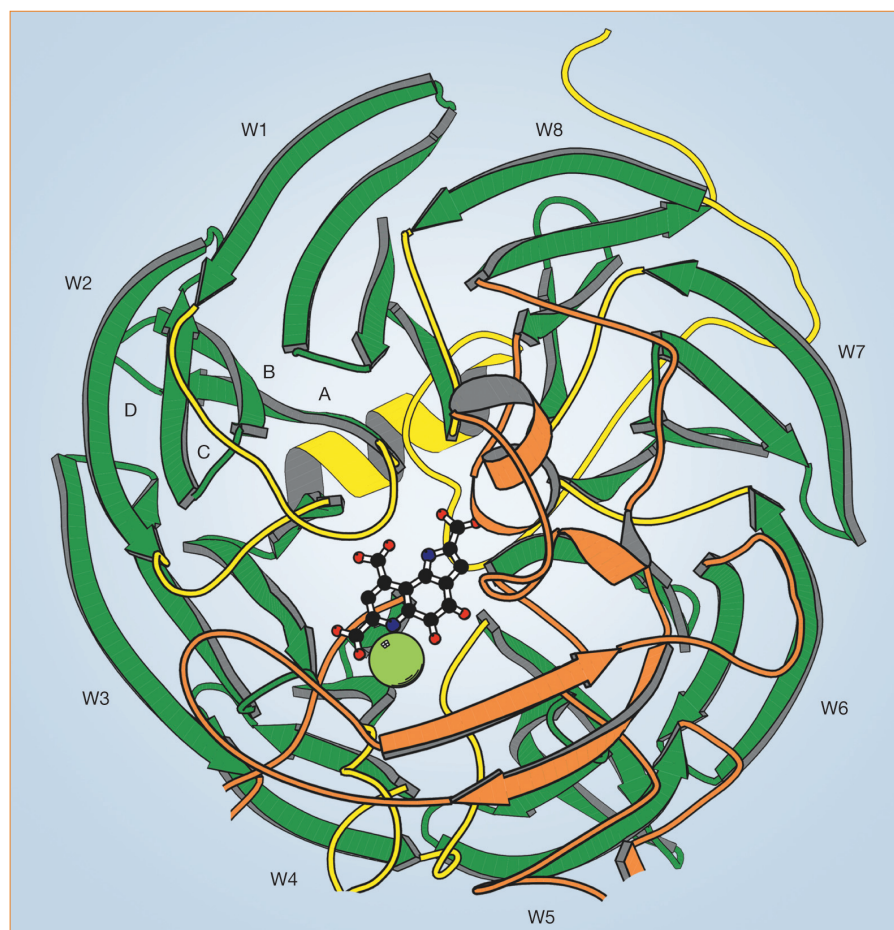


Figure 1 The superbarrel structure of the α -subunit of the PQQ-dependent methanol dehydrogenase⁶. This is simplified to show the β -sheets (labelled W1–W8) that form the 'propeller blades' of the propeller fold; each blade is made up of four antiparallel strands (labelled A–D). It can be seen that these eight β -sheets have no specific role in binding the PQQ molecule (shown as ball-and-stick model). The calcium ion is shown as a small sphere.

Table 1 Dietary PQQ status and plasma AAA-to-lysine ratio in rodents

	Dietary lysine (% by weight)		PQQ (nmol g ⁻¹ diet)	Plasma AAA (nmol ml ⁻¹)		Plasma lysine (nmol ml ⁻¹)		Ratio	
Rats	1.2		0	11.0		526		0.021	
			6.6	7.5		495		0.015	
Mice	Experiment			Experiment		Experiment		Experiment	
	A	B		A	B	A	B	A	B
	1.2	1.0	0	BD	BD	356	303	-	-
	1.2	1.0	6.6	BD	BD	362	354	-	-
	2.4	-	0	27	-	367	-	0.07	-
	2.4	-	6.6	29	-	423	-	0.09	-
	4.8	4.0	0	148	182	785	775	0.19	0.23
	4.8	4.0	6.6	86	76	564	757	0.15	0.10
	7.2	8.0	0	136	132	793	789	0.17	0.16
	7.2	8.0	6.6	87	89	578	794	0.15	0.11

Samples were taken by cardiac puncture at 2–4 h following initiation of the light phase of the dark/light cycle (12 h/12 h). Amino acids were determined using a Beckman 6300 amino-acid analyser and established procedures. The s.e.m. for each value ranged from 8–14%; the number of rats or mice was 6–14 for each determination. BD, below the detection limit.

the rates of AAS oxidation in cytosolic fractions of liver extract (from 5-week-old rats) ranged from 20 to 40 pmol min⁻¹ mg⁻¹, whether PQQ was added to (6 nmol g⁻¹ of diet) or absent from their diet. Moreover, plasma AAA concentrations were not altered significantly by PQQ status in ways that could be ascribed to reduced AAS dehydrogenase activity (Table 1; see ref. 3 for diet and animal husbandry details).

In rats and mice, the ratio of amino adipic acid to lysine tended to be slightly greater in PQQ-deprived animals. In rats, this ratio was inversely associated with body weight: that is, the concentration of amino adipic acid (in nmol ml⁻¹) was 28.100–0.086 times the body weight in g; the correlation coefficient r^2 was about 0.48. In addition, when C57 mice were exposed to varying amounts of dietary lysine to induce enzymes related to lysine catabolism, according to the rationale of Kasahara and Kato¹, we found that there was no significant change in the ratio of serum AAA to lysine at any given level of dietary lysine or for different intakes of PQQ.

Kasahara and Kato point out that the principal pathway for lysine degradation, the saccharopine pathway, is mitochondrial. However, the method we used to obtain tissue extracts as a source of AAS dehydrogenase did not differ significantly from another method⁵ used to obtain extracts for assaying the activity of lysine oxoglutarate reductase and saccharopine dehydrogenase, the principal enzymes in the so-called saccharopine pathway for lysine degradation; these two enzymes seem to be associated with the mitochondrial matrix⁶ and a related AAS dehydrogenase activity. In this regard, the exact cellular localization of AAS dehydrogenase remains unknown.

Our observations of no apparent change in plasma amino adipic acid levels in response to changes in PQQ and lysine nutritional status are important. The saccharopine–lysine degradation pathway is induced during starvation or when dietary lysine is in excess^{6,7}. At about 1% lysine in the diet (the requirement for rodents), the plasma concentrations of

amino adipic acid are normally low, but increase in response to increased dietary lysine. We have found that PQQ deficiency does not influence the AAA/lysine ratio in three separate experiments (Table 1).

In our early work², we reported that PQQ deficiency was associated with a decrease in lysyl oxidase, which was thought at the time to require PQQ. The association between lysyl oxidase and PQQ was based on Raman spectral data and inferences were made largely from spectroscopic studies. The inferences in this case were incorrect and lysine tyrosylquinone was later identified as the cofactor. We now know that our observations of decreased lysyl oxidase activity in PQQ-deprived mice were due mostly to diminished expression of lysyl oxidase as a result of impaired growth³. Although rapid progress is often made from processes that take advantage of strong inference, there are pitfalls. At this point, we persist in our view that PQQ may be an important biological factor for which nutritional and pharmacological exposure can elicit positive and presumed healthy outcomes. However, we suggest that insufficient information is available so far to state that PQQ uniquely performs an essential vitamin function in animals.

Robert Rucker*, **David Storms***,
Annemarie Sheets*, **Eskouhie Tchapanian***,
Andrea Fascetti†

Departments of *Nutrition, College of Agriculture and Environmental Sciences, and †Molecular Bioscience, School of Veterinary Medicine, University of California, Davis, California 95616, USA

e-mail: rbrucker@ucdavis.edu

doi:10.1038/nature03323

1. Kasahara, T. & Kato, T. *Nature* **422**, 832 (2003).
2. Killgore, J. et al. *Science* **245**, 850–852 (1989).
3. Steinberg, F. M., Gershwin, M. E. & Rucker, R. B. *J. Nutr.* **124**, 744–753 (1994).
4. Chang, Y. F., Ghosh, P. & Rao, V. V. *Biochim. Biophys. Acta* **1038**, 300–305 (1990).
5. Papes, F., Kemper, E. L., Cord-Neto, G., Langone, F. & Arruda, P. *Biochem. J.* **344**, 555–563 (2002).
6. Blemings, K. P., Crenshaw, T. D., Swick, R. W. & Benevenga, N. J. *J. Nutr.* **124**, 1215–1221 (1994).
7. Wang, S. X. et al. *Science* **273**, 1078–1084 (1996).

Reply: T. Kasahara and T. Kato reply to this communication (doi:10.1038/nature03324).

Kasahara and Kato reply — Our announcement¹ that the redox cofactor pyrroloquinoline quinone (PQQ) is a new vitamin has been challenged as being premature by Felton and Anthony² and by Rucker et al.³.

Felton and Anthony² assert that the repeated sequences we find in mouse U26 (mU26) represent a β -propeller protein and not a PQQ-dependent enzyme. Anthony and a colleague claimed previously that the 'tryptophan docking motif', a partial sequence of the PQQ enzyme repeat, is present only in PQQ-dependent enzymes that have a different number of blades and a unique manner of closure compared with other β -propeller proteins⁴.

Our identified mU26 matches these profiles (Fig. 1a). Although any amino-acid residue in the PQQ enzyme repeat cannot easily bind to a PQQ molecule directly⁴, the β -propeller structure formed by the PQQ enzyme repeats does provide a space for PQQ. Also, the repeated sequences of mU26 are more similar to those of bacterial PQQ-dependent enzymes than to those of other β -propeller proteins such as G-protein β -subunits, Kelch or RCC1 proteins (Fig. 1b).

Rucker et al.³ find that the amino adipic 6-semialdehyde (AAS) dehydrogenase activity in their samples with and without PQQ does not change, nor does the ratio of plasma 2-amino adipic acid (AAA)/lysine between PQQ-deprived and PQQ-supplemented rodents³. However, they overlook the possibility that there may be two isozymes catalysing oxidation of AAS to AAA. There are two pathways for lysine degradation in mammals^{5,6}, and the isozyme function in each pathway might require a different redox cofactor. Because their experiments are based on earlier work⁷ in which the formation of NADH (the reduced form of nicotinamide adenine dinucleotide) from NAD⁺ is measured spectroscopically to determine enzyme activity, it might be impossible to detect PQQ-dependent enzyme activity.

As Rucker et al.³ do not measure plasma AAA concentrations under the same conditions as ours because of technical limitations, their results from PQQ-deficient rodents do not contradict ours. In addition, the data presented by Rucker et al. are undermined in the absence of appropriate statistical analysis. Under the most similar conditions (2.4% dietary lysine in experiment 1), their results seem, in fact, to support ours. Their data show that PQQ levels do affect lysine metabolism when a lysine-enriched diet is given. Lysine loading induces the expression of related enzymes involved in lysine-degradation pathways^{1,8}, which complicates data interpretation, but nonetheless supports PQQ's participation in lysine metabolism.

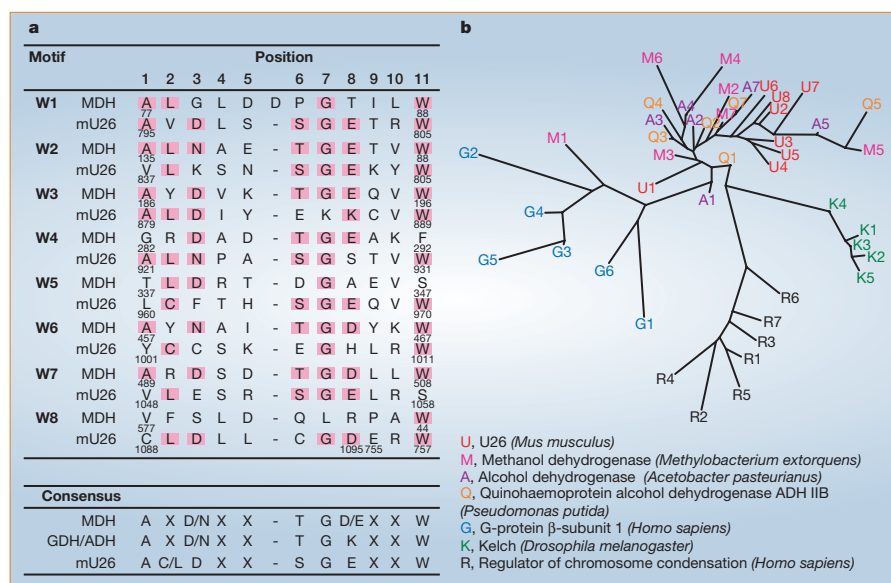


Figure 1 Characterization of the repeated sequences of murine U26 protein (mU26). **a**, Amino-acid sequences of the tryptophan docking motif of methanol dehydrogenase (MDH) from *Methylobacterium extorquens* and mU26 (modified from Fig. 4 of ref. 4). The amino acids, shown in pink, are identical to consensus sequences. In both MDH and mU26, β -propeller closure is achieved within the tryptophan docking motif (W8). GDH, glucose dehydrogenase; ADH, quinohaemoprotein alcohol dehydrogenase. **b**, Unrooted 'phylogenetic' tree of repeated sequences of β -propeller proteins constructed according to amino-acid similarity. As it is doubtful whether these sequences (PQQ enzyme repeats in U26 (U), methanol dehydrogenase (M), alcohol dehydrogenase (A) and quinohaemoprotein alcohol dehydrogenase (Q); WD repeats in G-protein β -subunit (G); Kelch repeats in Kelch (K); and RCC1 repeats in regulator of chromosome condensation (R)) evolved from a common ancestor sequence, this tree is not a real 'phylogenetic' tree. Branch lengths represent amino-acid differences: for example, M6 indicates the sixth repeated sequence in methanol dehydrogenase.

PQQ therefore seems to be an important nutrient for experimental rodents^{1,3,9,10}. Whether or not PQQ provides nutritional benefit for humans has yet to be determined. The dietary requirements of PQQ may vary between species in the same way as ascorbic acid (vitamin C), which is essential for humans but not mice. Developing a reliable method to microassay PQQ in human samples should help to resolve this issue.

Takaaki Kasahara, Tadafumi Kato

Laboratory for Molecular Dynamics of Mental Disorders, Brain Science Institute, RIKEN, Wako-shi, Saitama 351-0198, Japan

e-mail: kato@brain.riken.jp

doi:10.1038/nature03324

1. Kasahara, T. & Kato, T. *Nature* **422**, 832 (2003).
2. Felton, L. M. & Anthony, C. *Nature* doi:10.1038/nature03322 (2005).
3. Rucker, R., Storms, D., Sheets, A., Tchapanian, E. & Fascetti, A. *Nature* doi:10.1038/nature03323 (2005).
4. Anthony, C. & Ghosh, M. *Progr. Biophys. Mol. Biol.* **69**, 1–21 (1998).
5. Carson, N. A. J., Scally, B. G., Neill, D. W. & Carre, I. J. *Nature* **218**, 679 (1968).
6. Gatfield, P. D. *et al. Can. Med. Assoc. J.* **99**, 1215–1233 (1968).
7. Chang, Y.-F., Ghosh, P. & Rao, V. V. *Biochim. Biophys. Acta* **1038**, 300–305 (1990).
8. Blemings, K. P., Crenshaw, T. D., Swick, R. W. & Benevenga, N. J. *J. Nutr.* **124**, 1215–1221 (1994).
9. Killgore, J. *et al. Science* **245**, 850–852 (1989).
10. Steinberg, F. M., Gershwin, M. E. & Rucker, R. B. *J. Nutr.* **124**, 744–753 (1994).

Regulation of cap-dependent translation by eIF4E inhibitory proteins

Joel D. Richter¹ & Nahum Sonenberg²

¹Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

²Department of Biochemistry and McGill Cancer Centre, McGill University, Montreal, Quebec H3G 1Y6, Canada

Eukaryotic messenger RNAs contain a modified guanosine, termed a cap, at their 5' ends. Translation of mRNAs requires the binding of an initiation factor, eIF4E, to the cap structure. Here, we describe a family of proteins that through a shared sequence regulate cap-dependent translation. The biological importance of this translational regulation is immense, and affects such processes as cell growth, development, oncogenic transformation and perhaps even axon pathfinding and memory consolidation.

A single cell never exploits the full panoply of gene products available for its use; at any one time, the transcription of most genes is unwarranted and therefore these genes are silenced. Even for those mRNAs that are synthesized and transported to the cytoplasm, however, there are often other levels of regulation.

The past few years have witnessed an explosion in the number of mRNAs whose translation is recognized to be temporally and spatially regulated in various cell types. Although no single mechanism controls the translation of all mRNAs, emerging evidence indicates that the regulated binding of translation initiation factors (eIFs) to the 7-methyl guanosine residue that caps the 5' ends of all nuclear-encoded eukaryotic mRNAs is important. In particular, the interaction of the ribosomal-subunit-associated eIF4G with the cap-bound eIF4E is necessary for cap-dependent translation. A group of factors generically known as eIF4E inhibitory proteins modulate the eIF4G–eIF4E interaction. Whereas some eIF4E inhibitory proteins repress translation by associating with eIF4E on a large number of transcripts, others are tethered to specific subsets of mRNAs through interactions with RNA binding proteins, thus restricting their inhibition of translation to only certain mRNAs. Biologically, the eIF4E inhibitory proteins are enormously important; they control development and cell growth, repress tumour formation, and may influence critical neuronal events such as axon guidance and synaptic plasticity, a phenomenon that may underlie long-term memory storage.

Here, we present not only the molecular mechanisms by which some of these proteins control translation, but also describe a few of the biological processes they regulate. Although only a handful of these eIF4E inhibitory proteins have been identified, we suspect that there may be several others that await discovery.

The translation initiation machinery

Initiation is the rate-limiting step in translation and is the most common target of translational control. The mRNA 5' cap is bound by eIF4E, a heterotrimeric protein complex that is the focal point for initiation. eIF4G is the backbone of this complex; it interacts not only with eIF4E, but also with eIF4A, an RNA helicase that facilitates ribosome binding and its passage along the 5' untranslated region (UTR) towards the initiation codon. eIF4G also associates with eIF3, a multisubunit factor that bridges the proteins bound to the mRNA's 5' end with the 40S ribosomal subunit (Fig. 1). This ribosomal subunit comes 'pre-charged' as a ternary complex composed of eIF2, GTP and the initiator methionine-transfer RNA. With the aid of eIF4 initiation factor as well as ATP, this agglomeration of RNA and protein is thought to scan the mRNA in the 5' to 3' direction. When it encounters an AUG start codon in an optimal

context, other factors as well as the 60S ribosomal subunit are recruited and polypeptide chain elongation begins¹.

The eIF4E–eIF4G interface is an important target for translational control. The core portion of eIF4G that interacts with eIF4E is small—about 15 amino-acid residues². Strikingly, several other proteins contain similar peptide motifs, and it is this region that competes with eIF4G for binding to eIF4E; in this manner they control the rate of 40S ribosomal subunit association with mRNA, and hence translation initiation. A clear demonstration of why the competition between eIF4G and other proteins for interaction with eIF4E is so effective in preventing translation comes from X-ray crystallographic analysis. Peptides derived from the regions of eIF4G and an eIF4E inhibitory protein called 4E-BP (for 4E-binding proteins, also known as PHAS-I for phosphorylated heat and acid

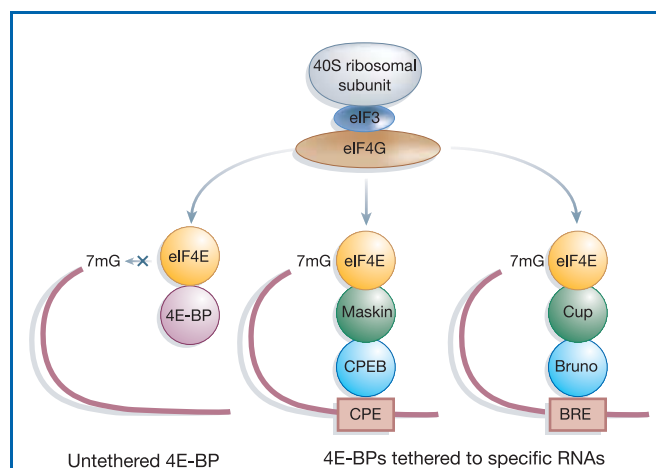


Figure 1 Translational control by eIF4E inhibitory proteins. Translation initiation occurs when the 40S ribosomal subunit is recruited to the 5' ends of capped (that is, 7mG-containing) mRNAs through an eIF3–eIF4G–eIF4E interaction. Initiation is disrupted by 4E-BP, which binds and sequesters eIF4E by interacting with eIF4G; this process occurs on a number of mRNAs because 4E-BP is not tethered to any particular sequence. Two examples of tethered 4E-BPs are represented by Maskin (*Xenopus*) and Cup (*Drosophila*). Through its association with CPEB, Maskin interacts with the eIF4E only on RNAs that contain a cytoplasmic polyadenylation element (CPE); disruption of the eIF4E–eIF4G complex by this protein is therefore mRNA-specific. In a similar manner, Cup, through its association with Bruno, binds and displaces the eIF4E only on mRNAs that contain a Bruno response element (BRE). Note, however, that Cup also binds Smaug, a protein that binds *nanos* mRNA; thus Cup associates with the eIF4E on this mRNA as well.

soluble protein stimulated by insulin) form nearly identical α -helical structures that lie along the same convex region of eIF4E, some distance from this protein's cap binding site^{3,4}. Peptides with the general sequence YXXXXL ϕ , where ϕ is any hydrophobic amino acid, would probably form similar α -helical structures, implying that other proteins containing this peptide motif could control translation initiation.

The original three eIF4E inhibitory proteins, the 4E-BPs, prevent eIF4F complex formation by sequestering available eIF4E (ref. 5). This sequestration results in the inhibition of translation of certain mRNAs that normally require high levels of available eIF4E (ref. 6). The newly discovered eIF4E-binding proteins described below interact with the eIF4E on only specific mRNAs, and do so either because they also interact with certain RNA elements directly, or do so through affiliations with RNA binding proteins.

eIF4E inhibitory proteins in development

One characteristic of early animal development is the synthesis and storage of mRNAs that are destined for later use. Although no single mechanism regulates the translation of all these mRNAs, one of the most important is the modulation of poly(A) tail length. For example, the 3' ends of most mRNAs terminate with ~150–200 adenylate residues that generally enhance stability and translation. In frog oocytes arrested at the end of meiotic prophase I, however, several dormant but stable mRNAs have relatively short poly(A) tails of approximately 20–40 bases. When the oocytes are stimulated to re-enter the meiotic divisions in preparation for fertilization, the poly(A) tails on these mRNAs are elongated to about 150 bases and translation ensues. Cytoplasmic polyadenylation is controlled by CPEB, a protein that interacts with the cytoplasmic polyadenylation element (CPE), a small sequence in the 3' UTR of mRNAs. CPEB also binds Maskin, a protein that interacts rather weakly with eIF4E probably because it contains an unusual threonine (T) for tyrosine (Y) substitution in its eIF4E-binding domain. In spite of this weak binding, Maskin disrupts eIF4E–eIF4G interactions; the CPEB–Maskin–eIF4E complex therefore inhibits the translation of CPE-containing mRNAs specifically⁷. When frog oocytes are induced to complete meiosis, CPEB stimulates poly(A) tail growth; the newly elongated poly(A) tail then is bound by poly(A) binding protein (PABP), which in turn interacts with eIF4G. PABP-bound eIF4G then displaces Maskin from eIF4E, thereby inducing translation⁸ (Fig. 2).

Complexes that are functionally equivalent to the CPEB–Maskin–eIF4E trimer have recently been uncovered in *Drosophila*, where the asymmetric distribution of mRNAs and proteins in the egg determine the body plan in the embryo. In the anterior portion of the embryo, Bicoid, a homeobox-containing transcription factor, activates genes that control segmentation; in that region, Bicoid also represses translation of the mRNA encoding Caudal, a transcription factor that directs the formation of posterior structures. Bicoid not only binds a small sequence in the *caudal* mRNA 3' UTR (the Bicoid binding region or BBR), it is also retained on an affinity matrix composed of cap-bound eIF4E (ref. 9). Although the Bicoid used for these experiments was derived from a cell extract that could contain several cap binding proteins, it may have directly associated with eIF4E, possibly through a peptide motif that resembles those in the 4E-BPs and Maskin⁹. Through these associations, Bicoid may act as a composite of both CPEB and Maskin in that its binding to the BBR confers mRNA-specificity to possibly prevent eIF4E–eIF4G interactions, and thus repress translation.

Three recent studies have identified the protein Cup as a new Maskin-like molecule that controls germ-cell formation and axis specification in *Drosophila*. One substrate that is acted upon by Cup is *oskar* mRNA, which is synthesized in nurse cells and transported through cytoplasmic bridges (ring canals) to the oocyte, where it localizes to the posterior region and is translated. The protein Bruno is responsible for suppressing *oskar* mRNA translation during

transport, and does so not only through its association with the Bruno response element (BRE) in the 3' UTR, but also through an interaction with Cup, an eIF4E-binding protein¹⁰. The Cup–eIF4E interaction, like the CPEB–Maskin interaction, prevents assembly of the eIF4F complex and thereby inhibits translation. Cup may also function in correct mRNA localization through an additional interaction with Barentsz, a protein that binds the molecular motor kinesin¹¹.

A second RNA, whose expression is controlled by Cup, encodes Nanos, an RNA binding protein that represses *hunchback* mRNA translation in the posterior pole; this repression is necessary for the proper development of the abdomen and other posterior structures. Whereas some *nanos* mRNA is highly concentrated at the posterior pole, most of it is dispersed throughout the rest of the embryo. This dispersed *nanos* mRNA pool is translationally repressed, at least in part, by Smaug, a protein that binds a *cis* element in the *nanos* 3' UTR. Smaug also interacts with Cup, which as noted above binds eIF4E to the exclusion of eIF4G, repressing translation¹².

eIF4E inhibitory proteins in cancer

The eIF4E–4E-BP interaction has important implications for disease, including cancer. This is not surprising given that overexpression of eIF4E not only causes malignant conversion of rodent fibroblasts¹³ and human mammary epithelial cells¹⁴, but also promotes tumour formation in transgenic mice^{15,16}. Consistent with the oncogenic potential of eIF4E, several studies have shown that human tumours express abnormally high levels of this protein¹⁷. Because the 4E-BPs inhibit eIF4E activity, it would not be surprising if they acted as tumour suppressors. Indeed, ectopic expression of the 4E-BPs in transformed cells partially reverts their transformed phenotype¹⁸. Interestingly, in several transformed mammary cell lines the 4E-BPs are hyperphosphorylated, which results in their dissociation from eIF4E and, in effect, increases the active concentration of this initiation factor¹⁴. It is of particular interest that

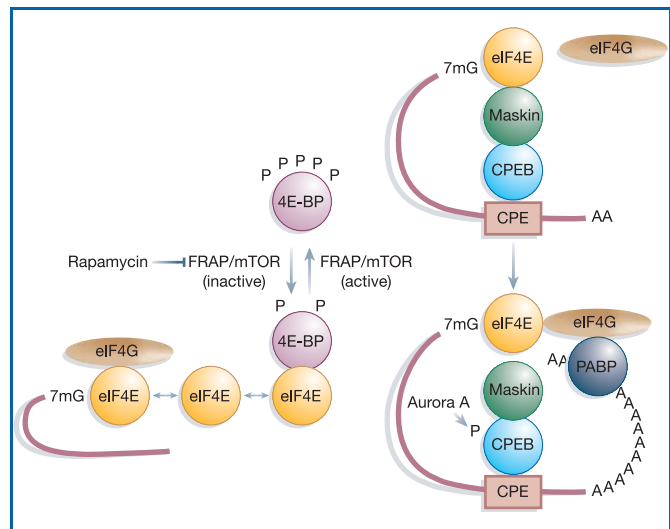


Figure 2 Regulation of eIF4E inhibitory proteins. The kinase FRAP/mTOR hyperphosphorylates 4E-BP on several sites; this causes the liberation of eIF4E from 4E-BP, and the association of eIF4E with both capped mRNA and eIF4G. The inhibition of FRAP/mTOR by rapamycin leads to the hypophosphorylation of 4E-BP and enhanced binding to eIF4E. Maskin binding to eIF4E excludes the eIF4G–eIF4E interaction on CPE-containing mRNAs. The inhibition of translation by Maskin is abrogated by cytoplasmic polyadenylation, which is induced by Aurora A-catalysed CPEB phosphorylation. The newly elongated poly(A) tail is bound by poly(A) binding protein, whose association with eIF4G helps disrupt the Maskin–eIF4E complex and facilitate initiation. The compartmentalization of Maskin- (or Cup-) bound eIF4E probably makes it resistant to further regulation by 4E-BP.

rapamycin, which specifically inhibits the kinase FRAP/mTOR and thus prevents 4E-BP hyperphosphorylation, is currently being tested as an anti-cancer agent¹⁹ (Fig. 2). Along the same lines, compounds that mimic the anti-translational activity of the 4E-BPs might also serve as useful agents to treat malignancies where eIF4E is abnormally high.

An increase in eIF4E amount or activity does not lead to elevated rates of global translation, but instead results in increased translation of a subset of mRNAs. These observations can be explained if cells contain an amount of eIF4E that is just enough to form sufficient eIF4F to maintain basal levels of translation²⁰ (but see also ref. 21). Because some mRNAs require more eIF4F owing to strong secondary structure in the 5' UTR, over-expression of eIF4E, which binds excess eIF4G and eIF4A (a factor that melts secondary structure), can selectively result in their increased translation²². This is clearly the case with ectopically expressed reporter mRNAs containing excessive secondary structure²³, and also with endogenous mRNAs encoding such proteins as Myc, fibroblast growth factor (FGF), ornithine decarboxylase (ODC) and vascular endothelial growth factor (VEGF)^{6,24}. These and other mRNAs play important roles in controlling cell growth and proliferation, and thus the marked effects of eIF4E on transformation could be explained by their elevated rates of translation.

eIF4E inhibitory proteins in the nervous system

Perhaps the newest frontier where translational control by the 4E-BPs is likely to have an enormous influence is the central nervous system. One neuronal activity that relies on protein synthesis is synaptic plasticity, which may be the underlying molecular and cellular basis for long-term memory consolidation. Synapses, the structures neurons use to communicate with one another, are plastic because they undergo biochemical and morphological modifications following their activation. These modifications are used by neurons to mould the strength of their responses. For these responses to endure, a neuron must establish a 'tag' at the stimulated synapse and then recognize and respond to it when the synapse undergoes subsequent stimulations²⁵. Although the nature of the tag(s) is not yet defined, its establishment or recognition requires protein synthesis, at or in close proximity to the synapse. Many different mRNAs are present in dendrites, but it is not clear how many are recruited for translation following synapse stimulation, nor is it certain which may be involved in synaptic tagging^{26,27}. Studies using rapamycin indicate that FRAP/mTOR signalling is involved in establishing the tag^{28,29}. Moreover, 4E-BPs are detected at synapses, suggesting that they may be the substrates of FRAP/mTOR signalling³⁰. Current studies using knockout mice should help to distinguish between the effects of rapamycin on 4E-BPs or other FRAP/mTOR substrates such as S6 kinase. In other studies, CPEB knockout mice have been found to have a deficit in some forms of synaptic plasticity³¹. These results, plus the observation that CPEB and Maskin immunoreactivity is detected at synapses^{32,33}, suggest that these molecules influence plasticity by controlling local CPE-dependent translation. However, whether mammals contain functional Maskin (or Cup) has not been demonstrated.

Finally, axons may rely on regulated mRNA translation to find their way to the appropriate destination. In *Xenopus*, protein synthesis inhibitors and rapamycin abrogate the attractive and repulsive turning of isolated retinal growth cones to netrin-1 and semaphorin 3A (ref. 34). These signalling agents also induce 4E-BP phosphorylation, suggesting that growth cone turning is mediated by cap-dependent translational control³⁴.

Outlook

By deciphering the mechanisms of translational control by eIF4E inhibitory proteins, new insights into how molecular decisions are made in normal and abnormal cells have been revealed. We have

focused on development, cancer and neurobiology, three areas in which translational control by these proteins clearly has a great impact, and where we think significant discoveries will yet be made. There has also been substantial progress in understanding how eIF4E inhibitory proteins influence cell-cycle progression and metabolism^{35–37}. Whereas recent studies report that several mammalian transcription factors can bind eIF4E (refs 38, 39), future experiments may uncover new eIF4E inhibitory proteins and demonstrate their importance in diverse biological processes. □

doi:10.1038/nature03205.

- Dever, T. Gene-specific regulation by general translation factors. *Cell* **198**, 545–556 (2002).
- Mader, S., Lee, H., Pause, A. & Sonenberg, N. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4γ and the translational repressors 4E-binding proteins. *Mol. Cell. Biol.* **15**, 4990–4997 (1995).
- Marcostrigiano, J., Gingras, A. C., Sonenberg, N. & Burley, S. K. Cocystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* **89**, 951–956 (1997).
- Matsuo, H. et al. Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. *Nature Struct. Biol.* **4**, 717–724 (1997).
- Haghighat, A., Mader, S., Pause, A. & Sonenberg, N. Repression of cap-dependent translation by 4E-binding protein 1: competition with p220 for binding to eukaryotic initiation factor-4E. *EMBO J.* **14**, 5701–5709 (1995).
- Gingras, A. C., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* **68**, 913–963 (1999).
- Stebbins-Boaz, B., Cao, Q., de Moor, C. H., Mendez, R. & Richter, J. D. Maskin is a CPEB-associated factor that transiently interacts with eIF-4E. *Mol. Cell* **4**, 1017–1027 (1999).
- Cao, Q. & Richter, J. D. Dissolution of the maskin-eIF4E complex by cytoplasmic polyadenylation and poly(A)-binding protein controls cyclin B1 mRNA translation and oocyte maturation. *EMBO J.* **21**, 3852–3862 (2002).
- Niessing, D., Blanke, S. & Jackle, H. Bicoid associates with the 5'-cap-bound complex of caudal mRNA and represses translation. *Genes Dev.* **16**, 2576–2582 (2002).
- Nakamura, A., Sato, K. & Hanyu-Nakamura, K. *Drosophila* cup is an eIF4E binding protein that associates with Bruno and regulates *oskar* mRNA translation in oogenesis. *Dev. Cell* **6**, 69–78 (2004).
- Wilhelm, J. E., Hilton, M., Amos, Q. & Henzel, W. J. Cup is an eIF4E binding protein required for both the translational repression of *oskar* and the recruitment of Barentsz. *J. Cell Biol.* **163**, 1197–1204 (2003).
- Nelson, M. R., Leidal, A. M. & Smibert, C. A. *Drosophila* Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. *EMBO J.* **23**, 150–159 (2004).
- Lazaris-Karatzas, A., Montine, K. S. & Sonenberg, N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature* **345**, 544–547 (1990).
- Avdulov, S. et al. Activation of translation complex eIF4F is essential for the genesis and maintenance of the malignant phenotype in human mammary epithelial cells. *Cancer Cell* **5**, 553–563 (2004).
- Ruggero, D. et al. The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis. *Nature Med.* **10**, 484–486 (2004).
- Wendel, H. G. et al. Survival signalling by Akt and eIF4E in oncogenesis and cancer therapy. *Nature* **428**, 332–337 (2004).
- Ruggero, D. & Pandolfi, P. P. Does the ribosome translate cancer? *Nature Rev. Cancer* **3**, 179–192 (2003).
- Rousseau, D., Gingras, A. C., Pause, A. & Sonenberg, N. The eIF4E-binding proteins 1 and 2 are negative regulators of cell growth. *Oncogene* **13**, 2415–2420 (1996).
- Sawyers, C. L. Will mTOR inhibitors make it as cancer drugs? *Cancer Cell* **4**, 343–384 (2003).
- Duncan, R., Milburn, S. C. & Hershey, J. W. Regulated phosphorylation and low abundance of HeLa cell initiation factor eIF-4F suggest a role in translational control. Heat shock effects on eIF-4F. *J. Biol. Chem.* **262**, 380–383 (1987).
- Rau, M., Ohlmann, T., Morley, S. J. & Pain, V. M. A reevaluation of the cap-binding protein, eIF4E, as a rate-limiting factor for initiation of translation in reticulocyte lysate. *J. Biol. Chem.* **271**, 8983–8990 (1996).
- Sonenberg, N. Translation factors are effectors of cell growth and tumorigenesis. *Curr. Opin. Cell Biol.* **5**, 955–960 (1993).
- Koromilas, A. E., Lazaris-Karatzas, A. & Sonenberg, N. mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. *EMBO J.* **11**, 4153–4158 (1992).
- Graff, J. R. & Zimmer, S. G. Translational control and metastatic progression: enhanced activity of the mRNA cap-binding protein eIF-4E selectively enhances translation of metastasis-related mRNAs. *Clin. Exp. Metastasis* **20**, 265–273 (2003).
- Martin, K. C. & Kosik, K. S. Synaptic tagging—who's it? *Nature Rev. Neurosci.* **3**, 813–820 (2002).
- Steward, O. & Schuman, E. M. Protein synthesis at synaptic sites on dendrites. *Annu. Rev. Neurosci.* **24**, 299–325 (2001).
- Richter, J. D. & Lorenz, L. J. Selective translation of mRNAs at synapses. *Curr. Opin. Neurobiol.* **12**, 300–304 (2002).
- Cassadio, A. et al. A transient, neuron-wide form of CREB-mediated long-term facilitation can be stabilized at specific synapses by local protein synthesis. *Cell* **99**, 221–237 (1999).
- Kelleher, R. J., Govindarajan, A., Jung, H., Kang, H. & Tonegawa, S. Translational control by MAPK signaling in long-term synaptic plasticity and memory. *Cell* **116**, 467–479 (2004).
- Tang, S. J. et al. A rapamycin-sensitive signaling pathway contributes to long-term synaptic plasticity in the hippocampus. *Proc. Natl Acad. Sci. USA* **99**, 467–472 (2002).
- Alarcon, J. M. et al. Selective modulation of some forms of Schaffer collateral-CA1 synaptic plasticity in mice with a disruption of the CPEB-1 gene. *Learn. Mem.* **11**, 318–327 (2004).
- Wu, L. et al. CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of α-CaMKII mRNA at synapses. *Neuron* **21**, 1129–1139 (1998).
- Huang, Y. S., Jung, M. Y., Sarkissian, M. & Richter, J. D. N-methyl-D-aspartate receptor signaling results in Aurora kinase-catalyzed CPEB phosphorylation and αCaMKII mRNA polyadenylation at synapses. *EMBO J.* **21**, 2139–2148 (2002).

34. Campbell, D. S. & Holt, C. E. Chemotropic responses of retinal growth cones mediated by rapid local protein synthesis and degradation. *Neuron* **32**, 1013–1026 (2001).
35. Pyronnet, S., Dostie, J. & Sonenberg, N. Suppression of cap-dependent translation in mitosis. *Genes Dev.* **15**, 2083–2093 (2001).
36. Heesom, K. J., Gampel, A., Mellor, H. & Denton, R. M. Cell cycle-dependent phosphorylation of the translational repressor eIF-4E binding protein-1 (4E-BP1). *Curr. Biol.* **11**, 1374–1379 (2001).
37. Tsukiyama-Kohara, K. *et al.* Adipose tissue reduction in mice lacking the translational inhibitor 4E-BP1. *Nature Med.* **7**, 1128–1132 (2001).
38. Topisirovic, I. *et al.* The proline-rich homeodomain protein, PRH, is a tissue-specific inhibitor of eIF4E-dependent cyclin D1 mRNA transport and growth. *EMBO J.* **22**, 689–703 (2003).
39. Nedelec, S. *et al.* Emx2 homeodomain transcription factor interacts with eukaryotic translation initiation factor 4E (eIF4E) in the axons of olfactory sensory neurons. *Proc. Natl Acad. Sci. USA* **101**, 10815–10820 (2004).

Acknowledgements Work in the laboratory of N.S. was supported by grants from the Canadian Institute of Health Research (CIHR), The National Cancer Institute of Canada and The Howard Hughes Medical Institute (HHMI) International Scholar Program. N.S. is a CIHR Distinguished Scientist and a HHMI International Scholar. Work in the laboratory of J.D.R. was supported by grants from the National Institutes of Health and the G. Harold and Leila Y. Mathers Charitable Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.D.R. (joel.richter@umassmed.edu).

Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*

Nicolas Gompel^{*†}, Benjamin Prud'homme^{*}, Patricia J. Wittkopp[†], Victoria A. Kassner & Sean B. Carroll

¹Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin, 1525 Linden Drive, Madison, Wisconsin 53706, USA

^{*} These authors contributed equally to this work

[†] Present addresses: Department of Zoology, Cambridge University, Downing Street, Cambridge CB2 3EJ, UK (N.G.); Department of Molecular Biology and Genetics, 227 Biotechnology Building, Cornell University, Ithaca, New York 14853, USA (P.J.W.)

The gain, loss or modification of morphological traits is generally associated with changes in gene regulation during development. However, the molecular bases underlying these evolutionary changes have remained elusive. Here we identify one of the molecular mechanisms that contributes to the evolutionary gain of a male-specific wing pigmentation spot in *Drosophila biarmipes*, a species closely related to *Drosophila melanogaster*. We show that the evolution of this spot involved modifications of an ancestral *cis*-regulatory element of the *yellow* pigmentation gene. This element has gained multiple binding sites for transcription factors that are deeply conserved components of the regulatory landscape controlling wing development, including the selector protein Engrailed. The evolutionary stability of components of regulatory landscapes, which can be co-opted by chance mutations in *cis*-regulatory elements, might explain the repeated evolution of similar morphological patterns, such as wing pigmentation patterns in flies.

The evolution of new morphological features is due predominantly to modifications of spatial patterns of gene expression. Changes in the expression of a particular gene can result from alterations either in its *cis*-regulatory sequences or in the deployment and function of the *trans*-acting transcription factors that control it, or both. Understanding the evolution of new morphological traits thus requires both the identification of genes that control trait formation and the elucidation of the *cis*- and *trans*-modifications that account for gene expression differences.

Evolution of *cis*-regulatory elements has been proposed to be a major source of morphological diversification because mutations in regulatory elements can produce discrete tissue-specific expression pattern changes while avoiding deleterious pleiotropic effects^{1–3}. In the best-studied cases of gene expression changes underlying morphological divergence, *cis*-regulatory modifications have been proposed^{4–6}, occasionally suggested by genetic evidence^{7–10}, but have only rarely been formally demonstrated¹¹ or analysed at the molecular level^{12,13}. It is currently not known whether the evolution of new morphological traits occurs largely through the modification of pre-existing *cis*-regulatory elements or from the generation of new elements; neither is it understood how many or what kinds of modifications are required for a regulatory element to drive a novel pattern.

To address these issues, we have analysed the evolution of a conspicuous male-specific wing pigmentation pattern in *Drosophila biarmipes*, a species closely related to *Drosophila melanogaster*¹⁴ (Fig. 1). Wing pigmentation patterns in insects are highly diversified and have various biological functions including mimicry, camouflage, thermoregulation, and mate selection¹⁵. In *D. biarmipes*, the sexually dimorphic wing pattern is associated with a courtship behaviour in which males display their wings conspicuously to the females, suggesting a function for this spot in mate choice^{16,17}. This wing spot has evolved recently in some species of the *D. melanogaster* group, such as *D. biarmipes*, and it is absent from close outgroup species such as *D. pseudoobscura*¹⁷ (Fig. 1).

Formation of wing pigmentation results from the conversion of melanin precursors diffusing from the veins into pigment deposits

at specific positions along the wing, wherever converting proteins are present¹⁸. The product of the *yellow* (*y*) gene is required for the production of black pigments, and the distribution of its product prefigures adult pigmentation patterns^{11,19}. The Yellow protein is expressed uniformly at low levels throughout the developing wings of *D. melanogaster* and *D. pseudoobscura*, where it imparts a low overall level of melanin pigmentation. In contrast, in *D. biarmipes*, in addition to the low, uniform expression, Yellow protein is highly expressed in an anterior distal spot¹⁹ (Fig. 1). This tight correlation between a novel Yellow expression pattern and a novel pigmentation

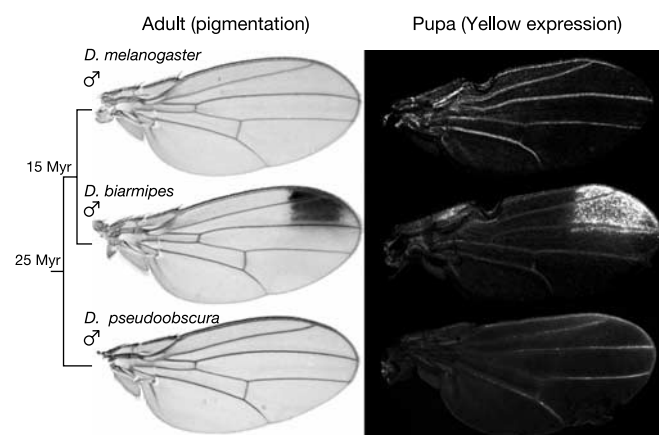


Figure 1 Expression of the Yellow protein prefigures adult wing pigmentation. The conspicuous spot of dark pigmentation present at the tip of the male wing of *Drosophila biarmipes* (left) is a new trait evolved among species of the *Drosophila melanogaster* group^{14,46} (about 15 Myr of divergence; divergence time is 60–80 Myr for the family Drosophilidae³⁰), superimposed on the ancestral pattern of uniform grey shading and darker veins found both in *D. melanogaster* and in *D. pseudoobscura*, a species from the sister *D. obscura* group (25 Myr of divergence^{29,30}). In all three species the male pupal distribution of Yellow in the wing, revealed by a specific antibody (right), foreshadows the adult pigmentation.

pattern prompted us to ask whether regulatory evolution at the *y* locus underlies the novel distribution of the Yellow protein in *D. biarmipes*, or whether this is due to changes in *trans*-acting regulators of *y*.

Regulatory changes in *cis* to the *yellow* locus

To test whether the observed differences in Yellow expression between *D. biarmipes* and *D. melanogaster* (Fig. 1) are due to changes at the *y* locus, we transformed *D. melanogaster* with green fluorescent protein (GFP)-reporter constructs containing non-coding DNA from the *D. biarmipes y* (*y^{bia}*) locus. If relevant evolutionary changes have occurred in *cis*, then the reporter gene might be regulated in *D. melanogaster* in a manner similar to the native *y* gene in *D. biarmipes*. If, however, the changes have occurred in *trans*, the *D. biarmipes y* regulatory element might drive reporter expression similar to that of the *y* gene in *D. melanogaster* (that is, uniformly). We found that *D. melanogaster* transgenic flies carrying the entire 5' region (8 kilobases; Fig. 2a) of the *y^{bia}* gene (5' *y^{bia}*) express GFP in the pupal wings in a pattern similar to the native *D. biarmipes* Yellow expression (Fig. 2b). Low levels of GFP are uniformly distributed across the wing, and higher levels of GFP are confined to the distal part of the anterior compartment. This result shows that the transcription factors deployed in the developing wing of *D. melanogaster* recognize *y^{bia}* *cis*-regulatory sequences.

Furthermore, the *D. biarmipes*-like expression pattern in a *D. melanogaster trans*-regulatory context shows that evolutionary changes in Yellow expression involve primarily *cis*-regulatory modifications at the *y* locus, which presumably entail the gain (or loss) of binding sites for transcription factors.

The 5' *y^{bia}* element does not recapitulate the precise restriction of the native spot of Yellow expression; higher levels of reporter protein expression extend along the proximal–distal axis, indicating that additional regulatory differences exist between *D. biarmipes* and *D. melanogaster*. Additional reporter constructs suggest that these differences are *trans* effects or are due to *cis*-acting elements located outside the region we have tested. The unique intron of the *D. biarmipes y* gene does contain another *cis*-regulatory element (for all developing sensory bristles) but has no activity in the wing other than in these sense organs. Furthermore, a transgene containing the 5' non-coding, 5' untranslated region, first exon, intron and second exon sequences (partial locus, Fig. 2a) is expressed in a similar pattern to that of the 5' *y^{bia}* element, indicating that the differences in *y* expression are not due to any of these sequences (data not shown).

Having localized major regulatory differences to the *y^{bia}* 5' region, we next investigated whether the novel *cis*-regulatory activity of the *y^{bia}* region arose in a pre-existing regulatory element or evolved *de novo* in the *D. biarmipes* lineage.

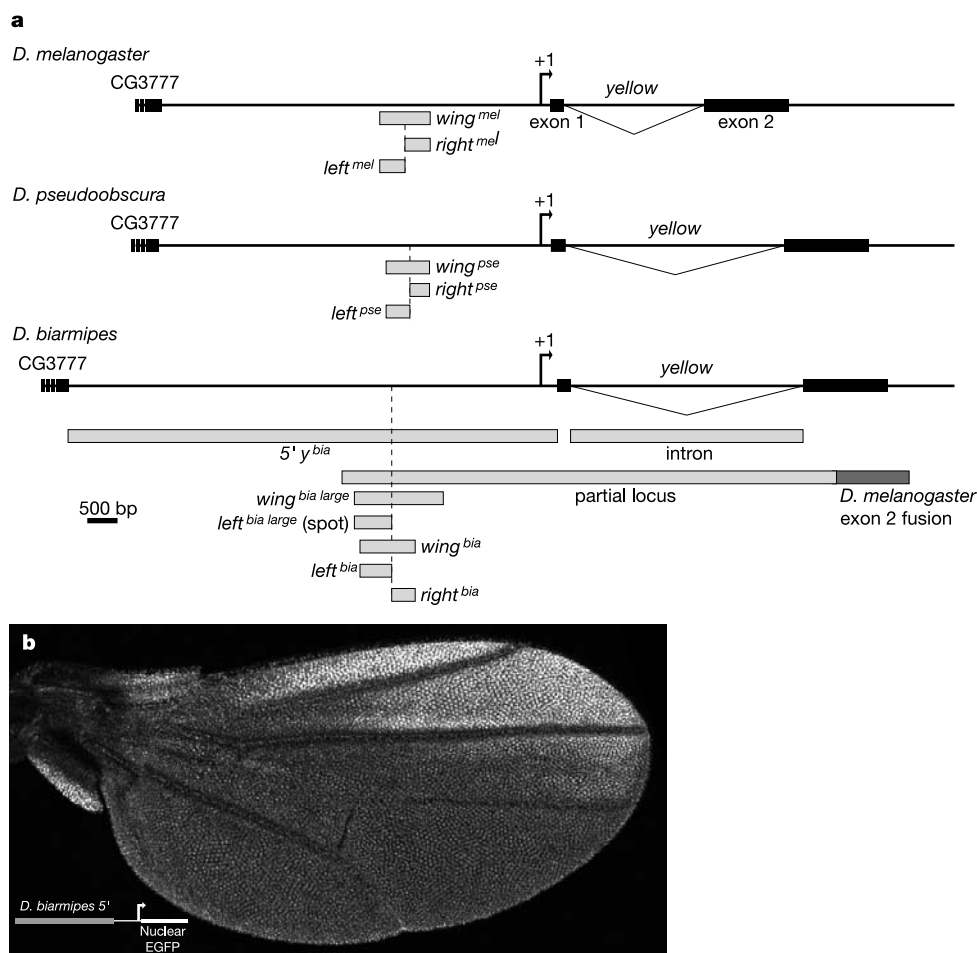


Figure 2 *Cis*-regulatory changes at the *yellow* locus are responsible for species-specific differences in Yellow distribution. **a**, The organization of the *y* locus is similar in *Drosophila melanogaster*, *Drosophila biarmipes* and *D. pseudoobscura*. **b**, The entire 5' region of *D. biarmipes y*, comprising sequences between the coding sequences of *y* and the closest predicted gene (CG3777), is sufficient to drive reporter GFP expression in *D. melanogaster*

at a time and in a pattern similar to those of *y* expression in native *D. biarmipes*. The *y^{bia}* intron does not drive wing expression other than in the marginal sensory bristles, and the partial locus drives expression in a pattern similar to the entire 5' region of *y^{bia}* (not shown). Black boxes, coding sequence; grey boxes, fragments analysed in transgenic constructs.

Evolution of a wing-specific *cis*-regulatory element

In *D. melanogaster*, analysis of the γ regulatory region has revealed that an 800-base-pair (bp) element located 1 kilobase upstream of the transcription start site, named the *wing^{mel}* element (Fig. 2a), is sufficient to drive gene expression throughout the pupal wing (Fig. 3b) and is necessary for adult wing pigmentation^{11,20,21}. We hypothesized that functional modifications of this element might account for differences in Yellow expression between the wings of *D. melanogaster* and *D. biarmipes*. There is strong sequence conservation of this portion of the γ locus between the two species (Fig. 3a and Supplementary Fig. 1). We transformed *D. melanogaster* with a GFP-reporter construct containing a 920-bp fragment from *D. biarmipes* orthologous to the *D. melanogaster* *wing* element (*wing^{mel}*), termed *wing^{bia}* (Fig. 2a). This fragment drives a reporter pattern resembling that driven by the 5' γ^{bia} element, with slightly less contrast between the levels of overall expression in the wing and in the anterior distal area (not shown). A larger fragment encompassing *wing^{bia}*, named *wing^{bia large}* (1,542 bp; Fig. 2a) drives a reporter pattern identical to the 5' γ^{bia} element (Fig. 3b). These results indicate that the sequences required for the strong anterior-distal activation of Yellow expression in *D. biarmipes* pupal wings are located within and immediately adjacent to a wing-specific *cis*-regulatory element that is orthologous to the *wing^{mel}* element.

To determine whether the novel *wing^{bia}* sequences evolved within an ancestral *wing cis*-regulatory element, we examined *D. pseudoobscura*, an outgroup species that belongs to a clade generally devoid of wing pigmentation patterns other than the grey (light black) homogeneous shading (Fig. 1). Phylogenetic character reconstruction suggests that the pigmentation spot was present in the common ancestor of *D. biarmipes* and *D. melanogaster* and has been lost in the *D. melanogaster* lineage^{22,23}. There is substantial sequence conserva-

tion at the γ gene between *D. biarmipes* and *D. pseudoobscura* (Fig. 3a and Supplementary Fig. 1), which allowed us to identify a region in *D. pseudoobscura* that is orthologous to the *wing^{bia}* element, named *wing^{pse}* (724 bp). This *wing^{pse}* element drives ubiquitous wing expression (Fig. 3b), demonstrating that a functional *wing* element is ancestral to the *D. melanogaster*/*D. biarmipes* lineage and that sequences within and/or adjacent to this element were modified to control high levels of expression in the anterior distal part of the wing in *D. biarmipes*.

To understand the organization of the *wing^{bia}* element and to localize its novel functional sequences, we further dissected the *wing^{bia}* element. We found that the sequences necessary for the anterior distal expression are separable from those controlling the general wing expression in *D. biarmipes*. Two complementary, non-overlapping sequences of the *wing^{bia}* element, *right^{bia}* and *left^{bia}* (Fig. 2a), drive respectively ubiquitous expression throughout the wing blade and strong activation in the anterior distal area of the wing (Fig. 3b) (as do the complementary subfragments from the *wing^{bia large}* element, *right^{bia large}* (not shown) and *left^{bia large}*; Fig. 4b). A similar dissection clearly separates two wing-specific complementary functions in *D. pseudoobscura* (ubiquitous expression, and expression around the veins) but yields non-functional elements in *D. melanogaster* (Fig. 3b). These results indicate that sites in both regions of the *wing* element are required for its function in *D. melanogaster* and *D. pseudoobscura*, and that some or all of the novel sequences in *D. biarmipes* responsible for the specific anterior distal wing expression of Yellow are located in the *left^{bia large}* element, hereafter referred to as the *spot* element. The distinct and robust activities of the two parts of the *wing^{bia}* element raise the possibility that the *wing* element has been subfunctionalized in *D. biarmipes* into two elements controlling expression throughout the wing and in the spot, respectively.

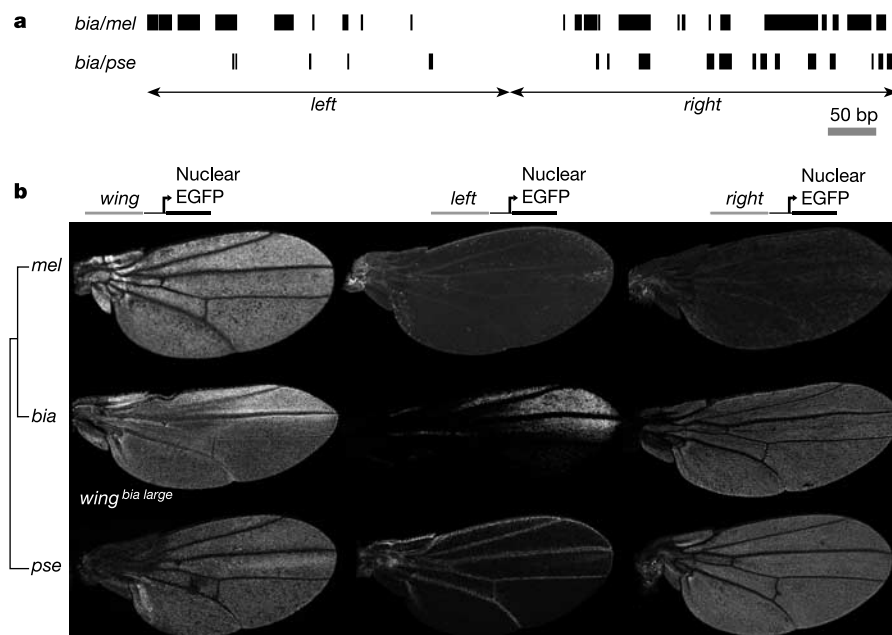


Figure 3 The *cis*-regulatory sequences governing spot formation evolved in the context of an ancestral wing enhancer. **a**, Conservation of the *wing* element sequence between *D. biarmipes* (*bia*) and *D. melanogaster* (*mel*) or *D. pseudoobscura* (*pse*) determined by Vista⁴⁷ with a 10-base-pair window length; only conservation above 75% is shown as solid boxes. Arrows show the boundaries of the *left* and *right* fragments. **b**, Reporter expression driven by the orthologous *wing* elements and its subfragments *left* and *right* (columns) of *D. melanogaster* (top), *D. biarmipes* (middle; the *wing^{bia large}* element is shown) and *D. pseudoobscura* (bottom), all expressed in *D. melanogaster*. The ubiquitous

expression driven by the outgroup species *wing^{pse}* element (expression is present in vein cells at a lower levels comparable to those in *left^{pse}*) shows that the sequences responsible for the spot pattern in *D. biarmipes* have evolved in the context of an ancestral wing regulatory element. The sequences controlling the spot pattern are separable from those controlling general expression in *D. biarmipes* (*left* and *right*). Note that the posterior boundary of activity of the *left^{bia}* construct lies near or at the anterior-posterior compartment boundary.

Multiple sites evolved in the wing spot element

In principle, the evolution of the spot pattern could arise through gaining binding sites for a single transcription factor that is expressed precisely in the cells that form the spot pattern. Alternatively, the spot pattern could result from the evolution of a combination of binding sites for multiple activators, as well as potential repressors that might restrict expression to this area. Resolution of these possibilities bears on the general question of

the number of steps involved in the evolution of new patterns of gene expression and *cis*-regulatory element function.

To distinguish between these possibilities, we derived a series of reporter gene constructs with smaller portions of the 675-bp *spot* element. A 196-bp construct (335–530; Fig. 4a) retained activity in the anterior distal region of the wing, although we noted that reporter expression now extended into the posterior compartment (Fig. 4d). This suggested that one or more sites critical for activation

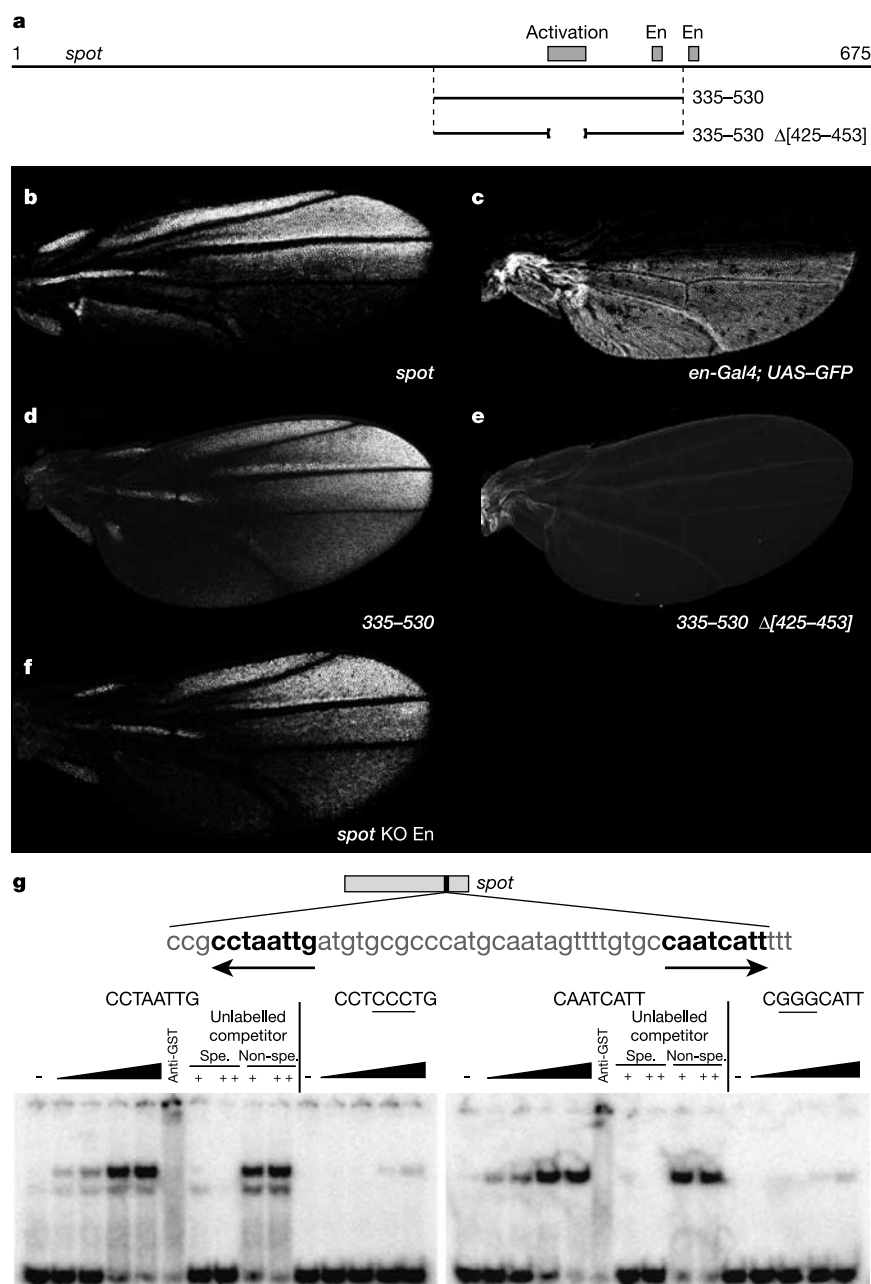


Figure 4 The *spot* element evolved through the acquisition of sites for both activators and repressors. **a**, Schematic of the 675-base-pair *spot* element showing the boundaries of deletion constructs and the location of identified binding sites. **b**, **d–f**, Expression of GFP driven by the *spot* element (**b**) and related constructs. **c**, The anterior border of expression of the selector gene *engrailed* abuts the *spot* element expression domain. **d**, A 196-base-pair element drives the *spot* pattern but is derepressed in the posterior compartment. **e**, Deletion of bp 425–453 abolishes activity of the *spot* element, indicating that sites required for activation lie within this region. **f**, Disruption of two characterized Engrailed binding sites from the *spot* element derepresses reporter expression in the posterior

compartment. **g**, The two candidate sites are bound specifically by the Engrailed protein *in vitro*. Increasing amounts of Engrailed homeodomain–GST fusion protein (0.25–5 nM) specifically shift labelled DNA oligonucleotides representing native sequences containing putative binding sites (left part of each gel) but not sequences in which Engrailed sites have been mutated (underlined in the sequences, right part of each gel). Addition of anti-GST antibody supershifts complexes. Addition of specific (spe.) or non-specific (non-spe.) unlabelled competitor DNA (+, 50 ng; ++, 500 ng) reveals the specificity of the formation of complexes. Supershift and competition experiments were performed in the presence of 5.0 nM protein.

resided in this 196-bp element and that one or more sites necessary for the restriction of expression from the posterior compartment resided outside it.

To localize further sequences required for activation of the *spot* element, we constructed a series of small deletions spanning the length of the 196-bp element (Fig. 4a, e, and data not shown). We found that a fragment lacking the internal sequences from bp 425 to 453 completely lacked reporter expression (Fig. 4e). This indicates that sequences required for activation in the *spot* are located within or overlap with bp 425–453. Together, these results indicate that sites for both at least one activator and one repressor have evolved in the *spot* element.

Direct regulation of the *spot* element by Engrailed

We next sought to identify potential *trans*-acting factors that regulate the *spot* element in *D. biarmipes*. The conspicuous posterior boundary of gene expression observed with the *left^{bia}* and *spot* elements (Figs 3b and 4b) is reminiscent of the compartment boundary of the wing²⁴ defined by the anterior border of expression of the selector transcription factor Engrailed²⁵ (Fig. 4c). The posterior expansion of GFP expression in the deletion constructs shown in Fig. 4d would be consistent with posterior repression of the *spot* element by Engrailed.

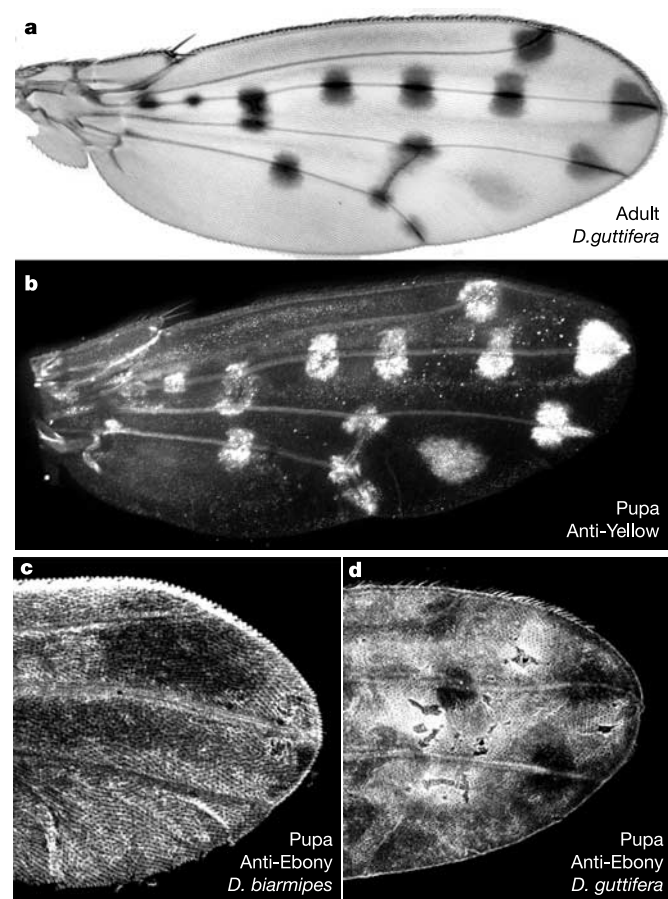


Figure 5 Concerted changes in the expression of Yellow and Ebony underlie the evolution of novel wing patterns. **a**, The distant species *D. guttifera* (a member of the *D. quinaria* group⁴⁸) has evolved a complex pattern of dark spots located at the intersection of wing veins and where campaniform sensilla form. The grey shading is also reinforced in some interveins. **b**, The pupal distribution of Yellow also prefigures this adult pattern. **c**, In *Drosophila biarmipes*, the spatial repression of Ebony is also associated with the formation of the adult male spot of pigmentation¹⁹. **d**, This repression of Ebony associated with pigmentation patterns seems to be general, because it is also seen in *D. guttifera*, where the adult spots will form.

To test whether Engrailed might be a direct regulator of the *wing^{bia}* element, we searched the *spot* sequence for putative Engrailed binding sites²⁶ and identified several candidate sites. Two of these sites, clustered within 43 bp, are specific to the *D. biarmipes spot* element (absent from *D. melanogaster* and *D. pseudoobscura* elements; Supplementary Fig. 1) and one site is located outside the 196-bp construct that exhibits some reporter expression in the posterior compartment (Fig. 4d). Gel-shift experiments on the native and mutated versions of these two sites showed that Engrailed binds specifically to them *in vitro* (Fig. 4g). Disruptions of these two Engrailed binding sites in the context of the *spot* element result in the specific derepression of reporter gene expression in the posterior compartment (Fig. 4f). These results show that the selector protein Engrailed directly represses the expression of the *y* gene in the posterior compartment of the wing and is one of the inputs that shapes the contours of the wing spot in *D. biarmipes*.

Multistep and multigenic evolution of the *spot*

Although multiple *cis*-regulatory modifications at the *y* locus have produced a profound evolutionary change in Yellow protein expression, it is important to ascertain whether changes in this one gene are sufficient for the evolution of the physical trait or whether additional evolutionary events are required. We have found that changes at *y* alone are not sufficient to create a pigmentation spot.

D. melanogaster y mutants carrying the *D. biarmipes y* gene (Fig. 2a, partial locus) recover only their species-specific pigment patterns; no wing spot is generated (not shown). Additional loci must therefore be involved.

The formation of pigment patterns is a multigenic process, and evolution at other pigmentation loci could also contribute to pattern evolution^{18,27,28}. The male spot of *D. biarmipes* is also associated with the localized downregulation of the melanin-inhibiting product of the *ebony* (*e*) gene during wing development¹⁹ (Fig. 5c), in a pattern that is approximately the inverse of Yellow expression. This suggests that, at least, both the repression of *e* and the activation of *y* are necessary for the formation of a dark spot. Consistent with this hypothesis is the observation that in *D. melanogaster e* mutants carrying the *y^{bia}* partial locus transgene (Fig. 2a), a slight darkening is observed specifically in the anterior area of the wing where *yellow* is strongly expressed (data not shown). However, this darkening is not comparable to the intense pigmentation spot of *D. biarmipes*. Changes in the expression of other pigmentation genes must also be involved. Furthermore, we have not been able to test whether changes in the *trans*-acting regulatory network of *D. biarmipes* might also contribute to the unique patterns of gene expression in the area of the wing spot. Taken together, these results indicate that the evolution of the novel pigmentation pattern of *D. biarmipes* required changes at multiple loci.

To determine whether the inverse regulation of expression of *y* and *e* is a general mechanism for the evolution of novel wing pigmentation patterns, we examined the expression of these proteins in *D. guttifera*, a species that separated from the *D. melanogaster* lineage about 40 Myr ago²⁹. This species has independently evolved a strikingly different and more complex wing pigmentation pattern (Fig. 5a). We found that the pattern of expression of the two proteins also exhibits an inverse relationship with higher levels of Yellow (Fig. 5b) and lower levels of Ebony (Fig. 5d) in the pupal wing where the eventual adult pigmentation spots will form. This indicates that the evolution of both *y* and *e* expression is involved in the formation and evolution of novel wing pigmentation patterns in drosophilids.

Chance caught on the wing: novelty by co-option

In drosophilid flies, the shape of the wings and the pattern of

venation have not changed much over 60–80 Myr of evolution^{30,31}. Their development and patterning are largely understood in *D. melanogaster* and the regulatory proteins involved are conserved³². One such protein, the selector protein Engrailed, is a deeply conserved feature of the compartmental organization of arthropod segments and appendages. In the *Drosophila* wing, Engrailed is part of the regulatory circuit that sequentially organizes the patterning of the anterior–posterior axis³³. Here we have shown that the activity of this transcription factor has been co-opted to control a feature of the novel wing pigmentation pattern in *D. biarmipes* through the evolution of specific binding sites within, or in the immediate vicinity of, a wing-specific regulatory element of the *y* gene. Because the expression driven by the *spot* element is also spatially modulated in *D. melanogaster*, this indicates that other conserved components of the wing *trans*-regulatory landscape (that is, one or more activators) have similarly been co-opted by the evolution of binding sites within the *y* wing element.

These findings suggest a general means by which novel expression patterns and characters can arise (Fig. 6a). Specifically, the random mutation of ancestral *cis*-regulatory elements (including point and insertional mutations) generates potential binding sites. If and when these sites can be recognized by transcription factors expressed in cells in which the ancestral element is active, the pattern or level of gene expression may be modified (Fig. 6a), in a manner similar to the mechanism of gene co-option demonstrated by the vertebrate crystallin genes³⁴. The patterns of expression of the eligible transcription factors are initially cryptic with respect to the target gene or trait, but these cryptic ‘prepatterns’ are revealed once functional binding sites have evolved in target genes. In this sense, and in this example, evolution is precisely a matter, as Jacques

Monod put it, of ‘chance caught on the wing’^{35,36}.

This model has two specific implications for the evolution of novel wing patterns. First, it explains how the observed diversity of wing pigmentation patterns might result from combinations of the numerous transcription factors expressed in the wing. Each of these combinations might constitute a distinct prepattern for pigmentation genes such as *y* or *e*, provided that the corresponding binding sites evolve in the proper *cis*-regulatory context. For instance, some of the spots on the wing of *D. guttifera* surround the sensory organs located on the veins, which form at similar positions in most drosophilids. This raises the possibility that transcription factors involved in the positioning of these landmark organs have been co-opted to change *y* or *e* regulation in *D. guttifera*. Second, this model might explain the widespread repeated evolution of strikingly similar pigmentation patterns observed in distantly related species (for instance, pigmentation patterns similar to those studied here have evolved independently in other dipterans; Fig. 6b). The evolutionary stability of the *trans*-regulatory landscape in drosophilid wings, reflected by the strong conservation of the wing shape and venation pattern in the family, suggests that similar pigmentation patterns might arise in parallel through the repeated evolution of binding sites for the same transcription factors in *cis*-regulatory regions of pigmentation genes. □

Methods

Fly stock and maintenance

Flies were bred at 25 °C on Wheeler–Clayton³⁷ or cornmeal³⁸ medium. Constructs were transformed into *D. melanogaster* *yw* mutants as described previously^{39,40}. The CantonS strain was used as wild-type *D. melanogaster*. *Drosophila pseudoobscura*, *Drosophila biarmipes* and *Drosophila guttifera* stocks were obtained from the Tucson stock centre (stock numbers 14011-0121.94, 14023-0361.01 and 15130-1971.10, respectively). All mature *D. biarmipes* males of this stock exhibited the wing spot. The *en-Gal4* and *UAS-GFP* stocks were obtained from the Bloomington *Drosophila* stock centre.

Immunocytochemistry

Pupal wings (70 h after puparium formation), still attached to the fly, were allowed to unfold in water after removal of the pupal cuticle. Flies were transferred to phosphate-buffered saline (PBS), in which the wings were cut off with a razor blade. Wings were fixed flat for 15 min between a slide and a coverslip in 4% formaldehyde PBT (PBS containing 0.03% Triton X-100), transferred on ice to a scintillation vial in the fixing solution for a further 15 min, sonicated briefly in the fixative with a Branson 200 ultrasonic cleaner, fixed for a further 30 min, washed with PBT, blocked for 1 h in PBT containing 1% bovine serum albumin, stained with a rat anti-*yellow* or a rabbit anti-*ebony* primary antibody¹⁹ and revealed respectively with a fluorescein isothiocyanate (FITC)-conjugated anti-rat antibody or FITC-conjugated anti-rabbit IgG antibody (Jackson ImmunoResearch).

Cloning

The *D. biarmipes y* locus sequence was amplified by direct and inverse polymerase chain reaction (PCR; details are available from the authors on request). The entire 5' region was amplified by PCR with primers designed in the coding sequences of *y* and the closest gene upstream of *y* in *D. melanogaster* (CG3777; ref. 41). All *y* fragments for reporter constructs were cloned into a customized version of the P-based transformation vector⁴² from which one of the two gypsy insulators had been removed and a new polylinker had been added. Fragments from *D. melanogaster* and *D. pseudoobscura* were amplified by PCR from genomic DNA and specific primers designed using available genome sequences^{43,44} (see Supplementary Table 1 for primer sequences).

Biochemistry

The *D. melanogaster* Engrailed homeodomain sequence was cloned into the glutathione S-transferase (GST) gene fusion vector pGEX-3X (Amersham Bioscience). The GST fusion protein was purified by affinity chromatography⁴⁵. DNA probes for electrophoretic mobility-shift assays were double-stranded oligonucleotides labelled with ³²P by end-filling in at both ends with the Klenow fragment of DNA polymerase I. Single-stranded oligonucleotides were annealed at a final concentration of 0.1 μM in 10 mM Tris-HCl pH 7.5 containing 0.1 M NaCl and 1 mM EDTA. Sequences of the oligonucleotide pairs were as follows: native sequences, 5'-TTTCCGCCTAAATGATG-3' and 5'-TTTCATCAATAGGCCG-3', 5'-TTTGGCCAATCATTTT-3' and 5'-TTTAAATGATTTGGCA-3'; mutated versions, 5'-TTTCCGCCTccTGATG-3' and 5'-TTTCATCAAgggAGGCCG-3', TTTTGGCGggCATTTT-3' and 5'-TTTAAATGcccGGCA-3'. Labelled probes were purified with G50 Sephadex beads (Sigma) on chromatography columns (Bio-Rad). DNA-binding assays, competition experiments and gel migrations were performed with 10–15 fmol of labelled probes (about 10⁴ c.p.m.) following a published protocol²⁶; they were pre-run for 0.5 h and run for 1.5 h at 4 °C on 8% native polyacrylamide minigels in 0.5 × Tris/borate/EDTA buffer pH 8.3. Non-specific competitor consisted of herring sperm DNA (Sigma) and the specific competitor was as used elsewhere²⁶.

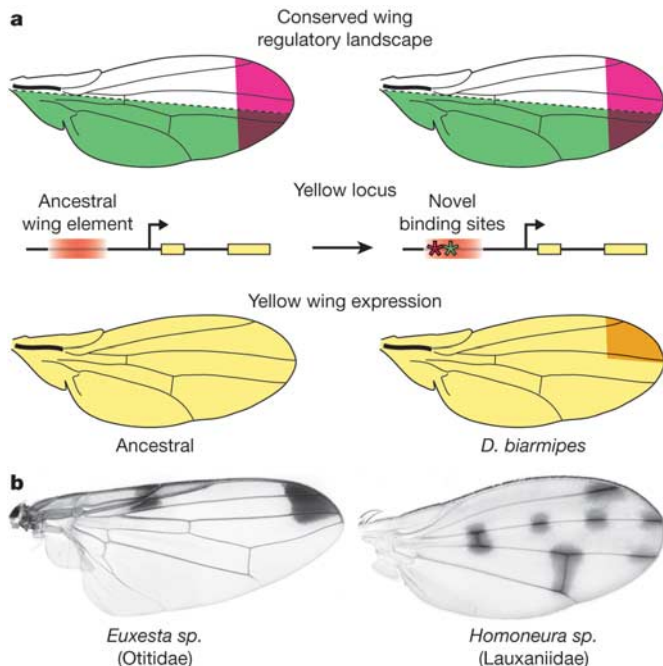


Figure 6 Cryptic prepatterns and the evolution of novel gene expression patterns through the evolution of *cis*-regulatory sequences. **a**, The upper panel shows a model of the conserved landscape of transcriptional regulators that pattern and shape the *Drosophila* wing (green and pink represent repressor and activator, respectively). The evolution of binding sites for a subset of these regulators in the *yellow* wing *cis*-regulatory element (coloured stars) co-opts them to modify *yellow* expression (lower panel). Combined with other regulatory changes at other loci, the changes at the *y* locus result in a novel pigmentation spot. **b**, Wing pigmentation patterns similar to *D. biarmipes* (left) or *D. guttifera* (right) evolved independently in other fly families (here Otitidae and Lauxaniidae).

Wing imaging

Adult wings were mounted flat in Hoyer's medium³⁸ and processed for bright-field imaging with a 4 × or 10 × dry lens on a Zeiss Axiophot microscope equipped with a Kontron charge-coupled device camera. For all reporter lines, pupal wings 70–90 h after puparium formation were mounted flat between a slide and a coverslip in PBT, without fixation, and imaged immediately with an Optiphot confocal microscope (Nikon) equipped with a 4 × dry lens and a BioRad 1024 system. Antibody-stained preparations were mounted in glycerol and imaged.

Received 30 September; accepted 1 December 2004; doi:10.1038/nature03235.

1. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* 2nd edn (Blackwell Science, Malden, Massachusetts, 2004).
2. Davidson, E. H. *Genomic Regulatory Systems: Development and Evolution* (Academic, San Diego, 2001).
3. Stern, D. L. Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
4. Averof, M. & Patel, N. H. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* **388**, 682–686 (1997).
5. Gompel, N. & Carroll, S. B. Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. *Nature* **424**, 931–935 (2003).
6. Yoon, H. S. & Baum, D. A. Transgenic study of parallelism in plant morphological evolution. *Proc. Natl Acad. Sci. USA* **101**, 6524–6529 (2004).
7. Sucena, E. & Stern, D. L. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. *Proc. Natl Acad. Sci. USA* **97**, 4530–4534 (2000).
8. Shapiro, M. D. *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).
9. Stern, D. L. A role of Ultrabithorax in morphological differences between *Drosophila* species. *Nature* **396**, 463–466 (1998).
10. Wang, R. L., Stec, A., Hey, J., Lukens, L. & Doebley, J. The limits of selection during maize domestication. *Nature* **398**, 236–239 (1999).
11. Wittkopp, P. J., Vaccaro, K. & Carroll, S. B. Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Curr. Biol.* **12**, 1547–1556 (2002).
12. Wang, X. & Chamberlin, H. M. Multiple regulatory changes contribute to the evolution of the *Caenorhabditis lin-48 ovo* gene. *Genes Dev.* **16**, 2345–2349 (2002).
13. Belting, H. G., Shashikant, C. S. & Ruddle, F. H. Modification of expression and cis-regulation of Hoxc8 in the evolution of diverged axial morphology. *Proc. Natl Acad. Sci. USA* **95**, 2355–2360 (1998).
14. Bock, I. R. & Wheeler, M. R. in *Studies in Genetics* (ed. Wheeler, M. R.) 1–102 (Univ. of Texas, Austin, 1972).
15. Majerus, M. E. N. *Melanism: Evolution in Action* (Oxford Univ. Press, Oxford, 1998).
16. Singh, B. N. & Chatterjee, S. Greater mating success of *Drosophila biarmipes* males possessing an apical dark black wing patch. *Ethology* **75**, 81–83 (1987).
17. Kopp, A. & True, J. R. Evolution of male sexual characters in the oriental *Drosophila melanogaster* species group. *Evol. Dev.* **4**, 278–291 (2002).
18. True, J. R., Edwards, K. A., Yamamoto, D. & Carroll, S. B. *Drosophila* wing melanin patterns form by vein-dependent elaboration of enzymatic prepatterns. *Curr. Biol.* **9**, 1382–1391 (1999).
19. Wittkopp, P. J., True, J. R. & Carroll, S. B. Reciprocal functions of the *Drosophila* Yellow and Ebony proteins in the development and evolution of pigment patterns. *Development* **129**, 1849–1858 (2002).
20. Chia, W. *et al.* Molecular analysis of the yellow locus of *Drosophila*. *EMBO J.* **5**, 3597–3605 (1986).
21. Geyer, P. K. & Corcos, V. G. Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in *Drosophila melanogaster*. *Genes Dev.* **1**, 996–1004 (1987).
22. Kopp, A. & True, J. R. Phylogeny of the oriental *Drosophila melanogaster* species group: a multilocus reconstruction. *Syst. Biol.* **51**, 786–805 (2002).
23. Schawaroch, V. Phylogeny of a paradigm lineage: the *Drosophila melanogaster* species group (Diptera: Drosophilidae). *Biol. J. Linn. Soc.* **76**, 21–37 (2002).
24. Garcia-Bellido, A., Ripoll, P. & Morata, G. Developmental compartmentalization of the wing disk of *Drosophila*. *Nat. New Biol.* **245**, 251–253 (1973).
25. Blair, S. S. Engrailed expression in the anterior lineage compartment of the developing wing blade of *Drosophila*. *Development* **115**, 21–33 (1992).
26. Solano, P. J. *et al.* Genome-wide identification of *in vivo* *Drosophila* Engrailed-binding DNA fragments and related target genes. *Development* **130**, 1243–1254 (2003).

27. Wittkopp, P. J., Williams, B. L., Selegue, J. E. & Carroll, S. B. *Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes. *Proc. Natl Acad. Sci. USA* **100**, 1808–1813 (2003).
28. Llopart, A., Elwyn, S., Lachaise, D. & Coyne, J. A. Genetics of a difference in pigmentation between *Drosophila yakuba* and *Drosophila santomea*. *Evolution* **56**, 2262–2277 (2002).
29. Russo, C. A., Takezaki, N. & Nei, M. Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.* **12**, 391–404 (1995).
30. Powell, J. R. *Progress and Prospects in Evolutionary Biology: the Drosophila Model* (Oxford Univ. Press, New York, 1997).
31. Hardy, D. E. *Diptera: Cyclorrhapha II, Series Schizophora, Section Acalypterae I, Family Drosophilidae* (Univ. of Hawaii Press, Honolulu, 1965).
32. De Celis, J. F. Pattern formation in the *Drosophila* wing: The development of the veins. *BioEssays* **25**, 443–451 (2003).
33. Blair, S. S. Compartments and appendage development in *Drosophila*. *BioEssays* **17**, 299–309 (1995).
34. Gonzalez, P., Rao, P. V., Nunez, S. B. & Zigler, J. S. Jr Evidence for independent recruitment of zeta-crystallin/quinone reductase (CRYZ) as a crystallin in camelids and hystricomorph rodents. *Mol. Biol. Evol.* **12**, 773–781 (1995).
35. Monod, J. *Le Hasard et la Nécessité. Essai sur la Philosophie Naturelle de la Biologie Moderne* (Éditions du Seuil, Paris, 1970).
36. Kauffman, S. A. *At Home in the Universe: the Search for the Laws of Self-organization and Complexity* (Oxford Univ. Press, New York, 1995).
37. Wheeler, M. R. & Clayton, F. E. A new *Drosophila* culture technique. *Drosophila Inf. Serv.* **40**, 98 (1965).
38. Ashburner, M. *Drosophila. A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
39. Spradling, A. C. & Rubin, G. M. Transposition of cloned P elements into *Drosophila* germ line chromosomes. *Science* **218**, 341–347 (1982).
40. Miller, D. E., Holtzman, S. L. & Kaufman, T. C. Customized microinjection glass capillary needles for P-element transformations in *Drosophila melanogaster*. *Biotechniques* **33**, 366–375 (2002).
41. Flybase. <http://flybase.bio.indiana.edu>.
42. Barolo, S., Carver, L. A. & Posakony, J. W. GFP and β-galactosidase transformation vectors for promoter/enhancer analysis in *Drosophila*. *Biotechniques* **29**, 726–732 (2000).
43. Celniker, S. E. *et al.* Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol.* **3**, RESEARCH0079 (2002).
44. Human Genome Sequencing Center. *Drosophila* genome project. <http://www.hgsc.bcm.tmc.edu/projects/drosophila/> (2002).
45. Schendel, P. F. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) 16.7.1–16.7.7 (Wiley, New York, 1993).
46. Bock, I. R. Current status of the *Drosophila melanogaster* species-group (Diptera). *Syst. Entomol.* **5**, 341–356 (1980).
47. Mayor, C. *et al.* VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**, 1046–1047 (2000).
48. Remsen, J. & O'Grady, P. Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. *Mol. Phylogenet. Evol.* **24**, 249–264 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. True, C. E. Nelson, C. M. Walsh and C. T. Hittinger for technical advice; J. True, S. Blair and members of the Carroll laboratory for discussions; B. L. Williams and J. Yoder for critical comments on the manuscript; S. Castrezana and T. Markow (Tucson *Drosophila* Stock Center) for providing *Drosophila* stocks; J. P. Gruber for the Euxesta sample; and S. Barolo for the pH Stinger vector. N.G. was funded by an EMBO long-term postdoctoral fellowship; B.P. and N.G. are recipients of a Philippe Foundation fellowship. The project was supported by the Howard Hughes Medical Institute (S.B.C.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.B.C. (sbcarrol@wisc.edu). The *D. biarmipes* y locus sequence is deposited in GenBank under accession number AY817623.

Structural basis of actin filament nucleation and processive capping by a formin homology 2 domain

Takanori Otomo, Diana R. Tomchick, Chinatsu Otomo, Sanjay C. Panchal, Mischa Machius & Michael K. Rosen

Department of Biochemistry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390, USA

The conserved formin homology 2 (FH2) domain nucleates actin filaments and remains bound to the barbed end of the growing filament. Here we report the crystal structure of the yeast Bni1p FH2 domain in complex with tetramethylrhodamine-actin. Each of the two structural units in the FH2 dimer binds two actins in an orientation similar to that in an actin filament, suggesting that this structure could function as a filament nucleus. Biochemical properties of heterodimeric FH2 mutants suggest that the wild-type protein equilibrates between two bound states at the barbed end: one permitting monomer binding and the other permitting monomer dissociation. Interconversion between these states allows processive barbed-end polymerization and depolymerization in the presence of bound FH2 domain. Kinetic and/or thermodynamic differences in the conformational and binding equilibria can explain the variable activity of different FH2 domains as well as the effects of the actin-binding protein profilin on FH2 function.

Actin nucleation is inherently slow, and cells have developed specialized machineries to accelerate this process. Formin proteins nucleate unbranched actin filaments that elongate from their fast-growing barbed ends^{1–3}. These structures have important roles in polarity, adhesion and cytokinesis in many cell types^{4–10}. Formins contain a conserved FH2 domain that mediates interactions with actin^{1,2}. The FH2 domain is often located carboxy-terminally to a proline-rich FH1 domain, which binds SH3 and WW domains, and profilin^{1,2}. FH2 domains from different formin proteins have a variety of actin regulatory activities. All can potentially nucleate new actin filaments^{5,11–15}, probably through stabilizing an actin dimer¹⁶. All FH2 domains also bind tightly to filament barbed ends (dissociation constant (K_d) $\sim$ 5–50 nM^{11,13–15,17}), with variable effects on filament dynamics. The FH2 domains of the actin regulatory proteins Bni1p, mDial and FRL potentially block the inhibitory activities of capping protein and gelsolin^{14–18}, but paradoxically only decrease rates of polymerization and depolymerization at the barbed end 2–5-fold, and have minor effects on critical concentration^{11,14–16,18}. These data, as well as real-time imaging of FH2-bound filament growth^{19–21}, suggest that these formins remain stably attached to the filament as actin monomers are inserted between the FH2 domain and the barbed end, leading to their description as leaky or processive caps¹⁸. The Cdc12p FH1FH2 element is apparently distinct and functions as a typical capping protein, which blocks barbed-end assembly and disassembly, and increases the filament critical concentration to that of the pointed end¹³. Formin activities can be markedly changed by profilin, which inhibits FH1FH2-mediated nucleation but relieves the cap in Cdc12p¹³, and can accelerate elongation rates of FH1FH2-bound barbed ends beyond that of free filaments²¹.

The FH2 domains of several formin proteins are dimeric in solution^{15,17}. The structures of an FH2 dimer from Bni1p and a truncated FH2 monomer from mDial were recently determined^{22,23}. These show a common elongated structural unit (referred to here as an actin bridge, see below), which in the FH2 dimer is reciprocally connected at both ends by flexible tethers to create a closed ring. The dimeric ring was proposed to bind to the two terminal actin monomers at the filament barbed end²², with each bridge engaging one actin. An alternating cycle involving release of one bridge and binding of a new actin monomer was

proposed as a mechanism for elongation of the FH2-bound barbed end^{18,22}.

We now describe the crystal structure of the FH2 domain of Bni1p in complex with actin, and propose a nucleating ratchet model for FH2 function. In the crystal, the bridge element binds two actin monomers in an orientation closely resembling a short-pitch actin filament, suggesting that this structure functions as a filament nucleus. Each pair of bridge elements contacts a total of three actin monomers, producing a structure that is free at the pointed end to bind the bulk filament, but sterically blocked at the barbed end. We propose that processive capping occurs through a dynamic equilibrium of the FH2 dimer at the barbed end between the blocked configuration, from which actin dissociation occurs, and an accessible configuration in which a free site on one bridge recruits new actin. Because the dynamics of the FH2 dimer are rectified by actin binding during filament growth, this model posits that formins can function as brownian ratchets. Differences in the actin binding and/or configurational equilibrium among different FH2 domains can account for their variable degrees of capping, as well as the effects of profilin on activity.

Structural overview

We determined the crystal structure of the FH2 domain of Bni1p (residues 1327–1769) in complex with tetramethylrhodamine-actin (TMR-actin, Table 1) using data to 3.05 Å. In the crystal, the actin molecules form a pseudo-paired filament along a 2₁ crystallographic screw axis. The FH2 domains wind around the outside of this structure in a helical arrangement (Fig. 1a).

The FH2 domain bound to actin is highly similar to the two previously reported structures of the free FH2 domain of Bni1p²² (Fig. 1b; see also Supplementary Fig. S1). It forms a shallow arc composed of a central, antiparallel, three-helix bundle flanked at its amino- and carboxy-terminal ends with two additional helical subdomains (termed knob and post, respectively²²). A “lasso”²² extends from the knob of one monomer and wraps around the post of its neighbour. The lasso of one monomer, and the post, three-helix bundle and knob of its partner in the FH2 dimer create the conserved structural unit. Because this unit bridges two actin monomers in our structure of the complex, we refer to it as the actin bridge element. In our crystal, actin is bound to ATP and adopts a

structure similar to that of an isolated TMR–actin–ATP complex²⁴. The DNase I binding loop of actin in the Bni1p complex is not observed in the electron density. In isolated TMR–actin, the covalently attached fluorophore is observed to be bound in the cleft between subdomains 1 and 3 at the barbed end of the monomer²⁴. In our crystals, there is only weak and fragmented electron density at the C-terminal end of actin (where TMR is covalently attached), suggesting that the fluorophore does not contribute strongly to the interactions between the FH2 domain and actin.

In the crystal, the lasso of each FH2 monomer interacts with the post of the next, resulting in a polymer of tethered bridges that winds around the pseudo-filament of actin (Fig. 1c). This arrangement is different compared with free Bni1p where two monomers form a ring-like dimer through reciprocal lasso–post interactions. There is no evidence that FH2 domains form polymers outside the crystal. Several FH2 domains, including those of Bni1p²², Dia (not shown) and FRL¹⁵, are dimeric in solution, and Bni1p, Cdc12p and mDia1 are functional when immobilized, where they are unlikely to polymerize. Thus, the FH2 domain seems to have undergone a domain swap during crystallization, exchanging one intra-dimer post contact for two inter-dimer contacts. We do not believe that this swap occurs during normal function of the FH2 domain. Rather, the biologically relevant form of the FH2–actin complex is, most likely, the ring-like structure formed by two reciprocally connected bridges. A model of the dimer can be readily generated from the structure of the polymer by swapping alternate lasso–post connections (Fig. 1c). In the polymer, the linker spans a distance of about 42 Å between the knob of one monomer (residue 1418) and the post/lasso of its partner (residue 1400). In the proposed functional dimer, this distance is about 34 Å, which could be spanned easily by the linker. Three successive actin monomers contact this reconstructed dimeric FH2 ring (Fig. 1c): an upper actin (3) binds the knob of one bridge (blue); a middle actin (2) binds the post/lasso of the first bridge and the knob of the second bridge (green); and a lower actin (1) binds the post of the second bridge.

The bridge creates a pseudo-short-pitch actin dimer

In the crystal, each bridge contacts two actin monomers through two distinct sites on opposite ends of the structure (Fig. 1b). One site is in the knob subdomain and contacts the barbed end of one actin monomer (Fig. 1b; see also Supplementary Fig. S2). The second is a composite surface formed by elements from both the lasso and post regions of the FH2 domain, which we will refer to as the post site (Fig. 1b; see also Supplementary Fig. S2). This surface binds subdomain 1 on the side of a second actin monomer, which is related to the knob-bound actin through the 2₁ screw axis of the crystal. Detailed descriptions of the knob and post interaction sites, as well as a large body of mutagenesis, NMR spectroscopy and sequence conservation data supporting the functional relevance of both sites, are provided in the Supplementary Methods.

The two actin monomers bound to the bridge element show remarkable similarity to the short-pitch dimer in the Holmes model of an actin filament (Fig. 2)^{25,26}. In both cases the actin monomers are oriented in parallel fashion with their N-terminal faces directed outward towards solvent/FH2, and their opposite faces inward. The actins are also radially displaced from the filament axis, so that subdomains 1 and 2 are more solvent-exposed than 3 and 4. In both cases, monomers are related by similar rotations and translations: 180° and 28 Å in our crystal, and 166° and 27.5 Å for the filament short-pitch dimer, although the axes for these operations are not identical (tilted by ~12°, Fig. 2). We will refer to the crystallographic organization as a pseudo-short-pitch actin dimer. Although the perfect 2₁ symmetry between the actins in the crystal is probably relaxed in solution, an ideal short-pitch filament orientation would require use of different surfaces on both actins and the FH2 domain from those observed in the crystal, or substantial changes in the geometry of the bridge element, or both (Supplementary Fig. S3). Thus, we believe that the pseudo-short-pitch geometry of the bridge-bound actins closely approximates an important functional state of the FH2 domain.

Three features of the bridge element–actin complex are particularly relevant to FH2 function. First, despite their similarities, the interaction surface in the bridge structure is much smaller than in

Table 1 Data collection, structure determination and refinement

Table 1. Data collection, structure determination and refinement

Data collection				
Crystal	Native	SeMet*	SeMet*	SeMet
Energy (eV)	12,425.4	12,681.2 (peak)	12,679.8 (inflection)	12,886.2 (remote)
Resolution range (Å)	31.78–3.05 (3.08–3.05)	39.20–3.50 (3.56–3.50)	39.20–3.50 (3.56–3.50)	39.20–3.50 (3.56–3.50)
Unique reflections	26,458 (1,314)	15,627 (675)	16,300 (719)	16,533 (635)
Multiplicity	2.0 (1.8)	7.3 (6.4)	7.1 (6.4)	3.6 (3.3)
Data completeness (%)	98.4 (90.8)	98.3 (89.6)	96.7 (82.5)	94.6 (74.4)
<i>R</i> _{merge} (%)†	4.6 (39.1)	7.0 (19.1)	6.7 (22.7)	6.1 (26.2)
<i>I</i> / <i>σ</i> (<i>I</i>)	24.4 (1.8)	31.3 (6.3)	29.7 (5.6)	18.2 (3.5)
Phase determination				
Anomalous scatterer	Selenium (six of eight possible sites)			
Figure of merit (35.0–3.50 Å)	0.36 (0.56 after density modification)			
Refinement statistics				
Resolution range (Å)	30.0–3.05			
Number of reflections <i>R</i> _{work} / <i>R</i> _{free}	22,738/1,145			
Wilson <i>B</i> -value (Å ²)	113			
Atoms (non-H protein/ATP/calcium)	6,164			
<i>R</i> _{work} (%)	28.9			
<i>R</i> _{free} (%)‡	31.3			
R.m.s.d. bond length (Å)	0.01			
R.m.s.d. bond angle (°)	1.78			
Cross-validated <i>σ</i> _A -coordinate error (Å)	1.16			
Mean <i>B</i> -value (Å ²)	103			
Missing residues	18 (1–3, 40–51, 373–375) for actin 32 (1327–1349, 1761–1769) for Bni1p			

Data for the outermost shell are given in parentheses.

* Bijvoet pairs were kept separate for data processing.

† $R_{\text{merge}} = 100 \sum_i \sum_j |I_{h,j} - \langle I_h \rangle| / \sum_i \sum_j I_{h,j}$, where the outer sum (*h*) is over the unique reflections and the inner sum (*j*) is over the set of independent observations of each unique reflection.

‡ *R*_{free} is calculated for a randomly selected 4.7%.

the filament model (470 compared with 1,700 Å²), as the monomers are further separated, and the finger is in a less-extended conformation than in the filament model (Supplementary Table S2). Thus, actin is probably in a higher energy (that is, strained) state when bound to the bridge than when incorporated into the filament. Second, the barbed end of the pseudo-short-pitch actin dimer is blocked by the bridge for addition of a third monomer in a filament-like orientation (Fig. 3a). An incoming monomer making perfect short-pitch contact to the actin monomer bound at the post site (yellow) would have a steric clash with the N termini of helices α I and α T of the bridge. A monomer making perfect long-pitch contact to the actin bound at the knob site (pink) would clash with the post site actin. An incoming monomer could interact in the crystal orientation, but actin-actin contacts in this configuration are minimal (see above). Finally, in contrast to the barbed end, the pointed end of the pseudo-short-pitch dimer is completely accessible to filament-like contacts to additional monomers (Fig. 3b). These features of the bridge-actin complex form a key foundation for our mechanistic models below.

Processive capping requires four actin-binding sites

In the crystal, each pair of bridge elements binds to a total of three

actin monomers (Fig. 1c). We propose that this arrangement represents an important configuration of the FH2 dimer at the filament barbed end. In this assembly, all four possible interaction sites of the FH2 dimer are engaged with actin, and all three terminal monomers are held in the strained orientation through bridge contacts. As in the single bridge complex described above, the actin 1/2 pair is blocked for monomer addition at the barbed end, whereas the actin 2/3 pair is accessible to pointed-end addition or interactions with bulk filament. This basic asymmetry could account for retention of the FH2 domain at the barbed end, rather than the pointed end or within the filament. In this complex there are only minimal contacts between monomer 1 and its neighbouring actins: the first bridge prevents filament-like short-pitch contacts to actin 2 (see above), and the geometry imparted by the pair of bridges prevents filament-like long-pitch contacts to monomer 3 (450 Å² buried surface versus 1,700 Å² in the Holmes model). Thus, monomer 1 is probably maintained at the barbed end largely through its contacts to the bridge rather than through contacts to the other actins.

To examine the importance of this four-site binding mode for dynamic aspects of FH2 function, we created a series of mutants with different numbers and combinations of available actin-binding

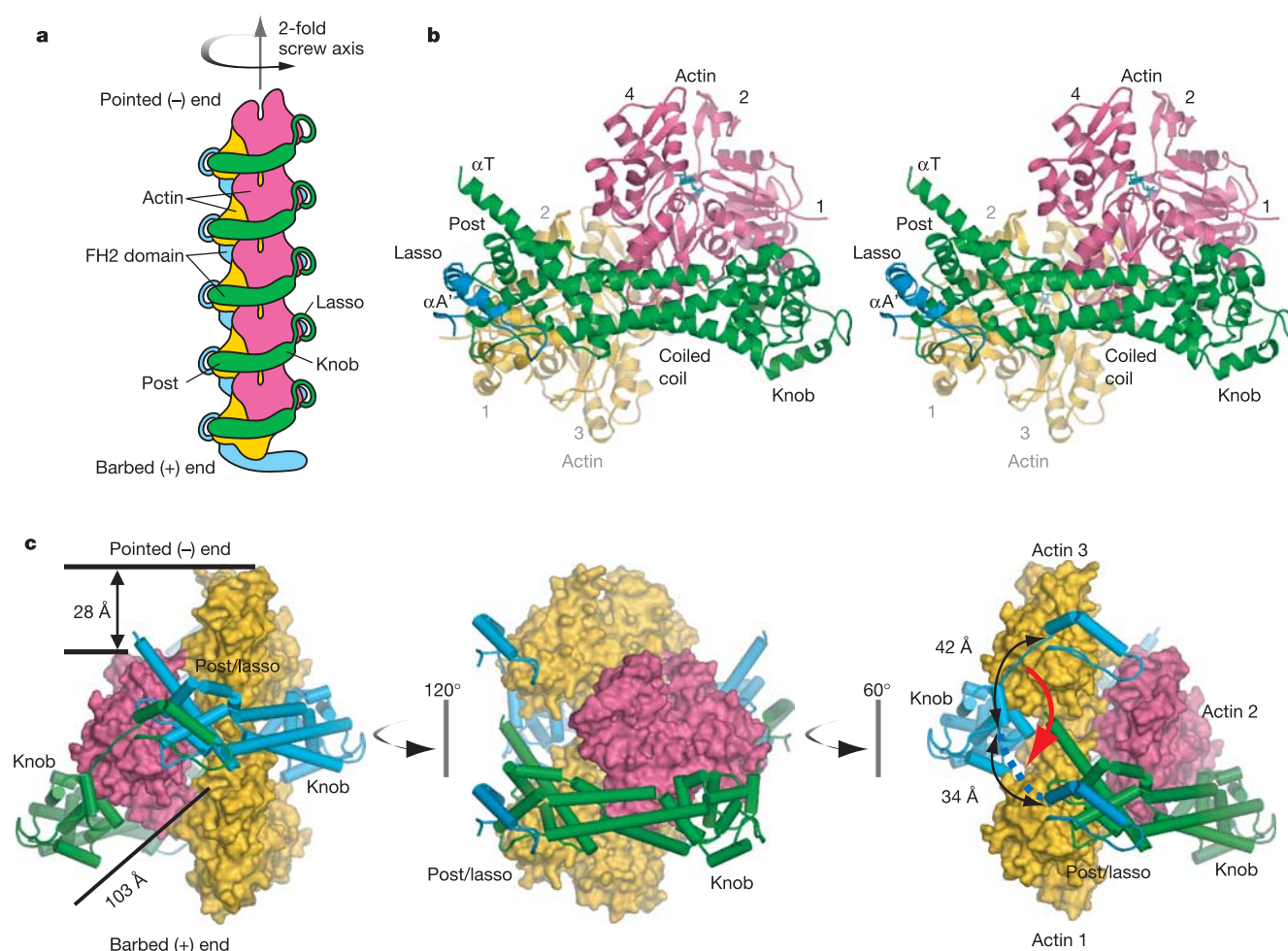


Figure 1 Structure of the FH2-actin complex. **a**, Cartoon representation of packing in the TMR-actin-FH2 crystal. The asymmetric unit contains one FH2 monomer shown in either green or blue and one actin shown in pink or yellow. **b**, Stereo view of the structure of one bridge element of the FH2 domain (blue and green) in complex with two crystallographically related TMR-actins (pink and yellow). Figures were made with PyMol³⁶. **c**, Three views of the structure of two bridge elements bound to three actins. Centre orientation is the same as in **b**. The FH2 monomers are shown as coloured

cylinders and TMR-actins in surface representation. The linker between lasso and knob in the green FH2 monomer can be swapped with a crystallographically related lasso (dark blue), as indicated by the red arrow, to generate the ring-like topology of the FH2 dimer (right panel). The other linker in the blue FH2 monomer remains as in the crystal (left panel). Two bridges in the TMR-actin complex are located farther apart (103 Å, left panel) than in the structure of the free Bni1p FH2 dimer (80 Å)²².

sites in the FH2 dimer. On the basis of the results with mutant homodimers (Supplementary Fig. S2; see also ref. 22), we used I1431A and K1601A mutations to disable the knob and post sites, respectively, on individual bridges in the FH2 dimer. We fused two FH2 domains together genetically using a linker. We refer to this protein as the linked wild type, or KP–KP, to indicate functional knob and post sites in both bridges. The resultant protein has the same size as the natural FH2 dimer, as measured by multi-angle light scattering (not shown), and behaves identically during purification and in actin assembly assays (Fig. 4; see also Supplementary Fig. S5), establishing that it reliably mimics the non-covalent dimer. We created two variants of this protein where a single knob or single post site was mutated (xP–KP and Kx–KP, respectively); two variants where one knob and one post site were mutated simultaneously in either the same bridge (xx–KP, *cis* KP mutant) or different bridges (xP–Kx, *trans* KP mutant); and the homodimeric double knob (xP–xP) and double post (Kx–Kx) variants (Supplementary Table S1).

At a concentration of 30 nM, the KP–KP protein and the wild-type FH2 dimer greatly accelerate filament assembly (Fig. 4a). Both single-site mutants are impaired in this assay, but retain activity that increases steadily to 500 nM, the maximum concentration tested (Supplementary Fig. S4). Of the four double mutants, only the xx–KP protein retains significant activity in this assay, whereas the others are virtually inactive up to concentrations as high as 500 nM (Supplementary Fig. S4). The activity of the xx–KP protein mirrors that reported previously for an isolated engineered bridge from Bni1p²². Thus, all the proteins possessing at least one intact bridge retained activity in the actin assembly assay, whereas those composed only of mutant bridges were inactive.

We also examined the effects of the mutants on barbed-end elongation (independent of nucleation) and filament depolymerization (Fig. 4b, c; see also Supplementary Fig. S5 and Table S1). The KP–KP protein and wild-type FH2 dimer each decrease the rate of both processes, with half-maximal inhibition at 3–9 nM. Elongation and depolymerization are maximally inhibited by ~30% and ~50%, respectively, exemplifying the leaky cap behaviour¹⁸. Although the single-site mutants (xP–KP and Kx–KP) had weaker activities in the actin assembly assay (Fig. 4a), both inhibited barbed-end elongation and depolymerization with high potency ($K_1 = 12\text{--}46\text{ nM}$) and to a much greater degree than the wild-type

linked protein (maximum elongation inhibition ~90%, depolymerization ~75%). Of the four double mutants, xx–KP decreased both rates to the level of the single-site mutants, but with a K_1 value of 170 nM, very similar to that reported for the isolated bridge²². The other three double mutants have only very weak effects on elongation and depolymerization. The ability of the single-site mutants to block barbed-end elongation yet still stimulate filament assembly (Fig. 4a) indicates that these proteins are potent nucleators of new filaments, but allow growth only at the pointed end.

Together, the data show that any proteins containing an intact bridge can nucleate new filaments. Those with only impaired bridges cannot (Fig. 4a), and appear to interact with filaments only weakly (Fig. 4b, c). All variants that can nucleate also probably support pointed-end growth (Fig. 4a–c). Thus, the bridge appears to be the minimal unit necessary and sufficient to promote filament nucleation. However, a single bridge blocks monomer addition and loss from the barbed end of existing filaments. Proteins with three intact sites bind filament barbed ends with higher affinity, but also have this same inhibitory function as the isolated bridge. Only the wild-type FH2 dimer composed of two bridges with four intact contact sites permits filaments to grow or shrink rapidly from the bound barbed end. This suggests that in the FH2 dimer, one bridge

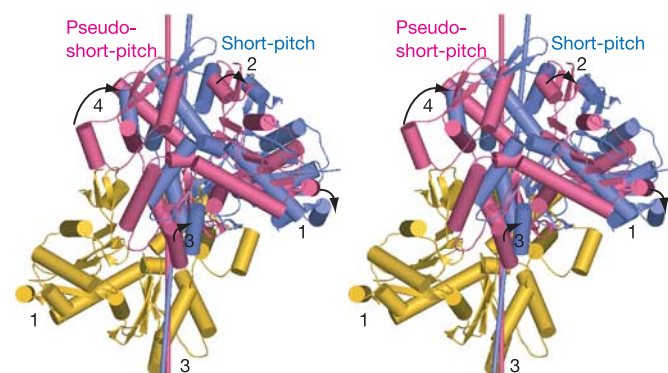


Figure 2 Relationship between the bridge-bound actin dimer and that in the Holmes model of the actin filament. The barbed-end monomers of each dimer were overlaid and only one is shown in yellow. The pointed end monomers of each dimer are shown in red (pseudo-short-pitch) and blue (filament-like short-pitch). The axes of the pseudo-filament (crystallographic 2_1 axis) and filament are shown in red and blue, respectively. The pointed-end monomers differ from each other by an ~2 Å translation perpendicular to the helix axis and rotation about the base of subdomain 3. Residues are displaced by <2 Å at the base of subdomain 3 to 5–15 Å at the tip of subdomains 1 and 4. In both arrangements, the finger loop (residues 262–274) of the lower actin contacts subdomain 3 of the upper actin^{25,26}.

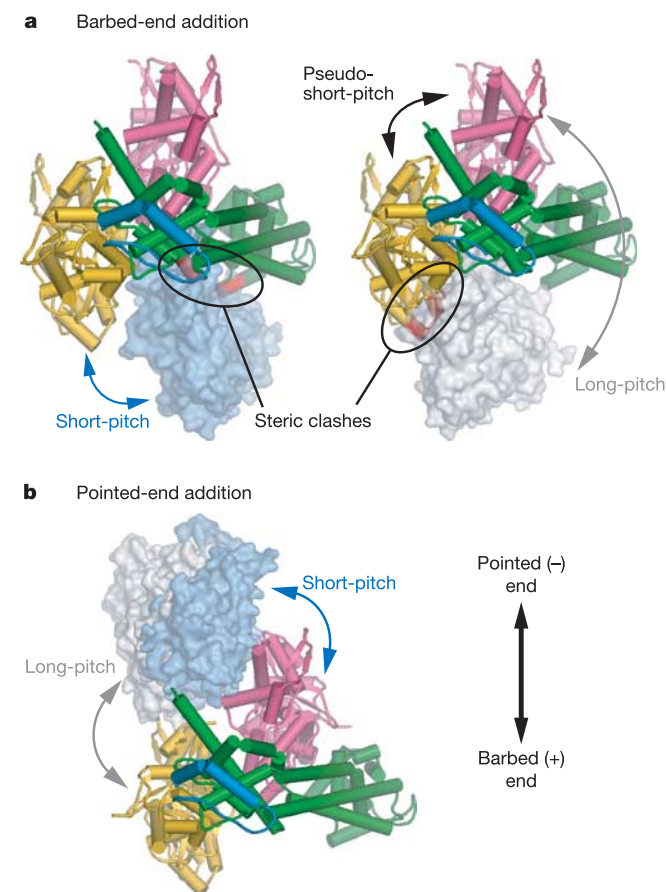


Figure 3 The bridge blocks actin addition at the barbed end but not the pointed end. The crystal structures of a bridge element and its two bound actins are shown in green/blue, pink and yellow cylinders, respectively. **a**, A third actin monomer adding at the barbed end is shown as a transparent surface modelled in filament-like orientation either short-pitch to the yellow actin (blue surface, left) or long-pitch to the pink actin (white surface, right). Structural elements from the bridge assembly that penetrate the surface of the incoming monomer are red. **b**, Actin monomers adding at the pointed end are shown as transparent surfaces with filament-like long-pitch orientation to the yellow actin (white surface) or short-pitch orientation to the red actin (blue surface). In both orientations, no steric clashes are observed with the bridge assembly.

blocks monomer binding and dissociation at the barbed end, and the second bridge provides functionality that overcomes this inhibition.

A nucleating ratchet model for FH2 function

Our structure of the bridge-bound actin and accompanying biochemical analyses suggest a straightforward mechanism for FH2-mediated filament nucleation. By stabilizing two actins in a conformation approximating the short-pitch helix, the FH2 domain could form a template from which a new filament could grow. Such a structure would overcome the largest energetic barrier to nucleation: formation of the initial actin dimer. Because profilin cannot be sterically accommodated at both positions in the dimer (not

shown), it should block FH2-mediated nucleation, as observed experimentally^{13–16}. This model requires that proteins contain at least one functional bridge for efficient nucleation, consistent with the lack of activity in the homodimeric mutants (Kx–Kx and xP–xP) and the *trans* double mutant (xP–Kx). Similarly, proteins with one wild-type bridge and additional functional sites in the second, such as the single-site mutants (Kx–KP and xP–KP) and the wild-type protein, should nucleate with greater potency than the xx–KP mutant, again as we observe.

Addition of actin monomer to this nucleus would then create the three-actin/two-bridge structure observed in the crystal. This assembly could grow (albeit slowly) in the pointed-end direction (Fig. 5, left panel), where it is unobstructed, and a new subunit (actin 4) could make ideal long- or short-pitch contacts to actins 2 or 3, respectively. Because both sets of idealized contacts cannot be made simultaneously, it is likely that the most favourable orientation of actin 4 is a variant of the filament structure that optimizes the total interaction energy to actins 2 and 3. Ideal contacts between actins 2 and 4 would position the long-pitch interface immediately adjacent to the finger loop of actin 3, providing a potential means of cooperatively stabilizing the incoming monomer, similar to the inter-strand interactions in free filaments^{25,26}. Further growth (actins 5, 6 and so on) would require similarly non-ideal contacts to actin 3, and eventually lead to an FH2–actin assembly attached to bulk filament. The transition from bridge-bound actin geometry to bulk filament geometry could occur entirely at the junction with bulk filament (as implied for simplicity in Fig. 5) or more gradually over a longer length. Recent evidence that actin filaments can have quite variable structures^{27,28} implies that such deviations from the idealized contacts of the Holmes model could be readily tolerated.

The FH2 domain nucleates filaments that grow rapidly in the barbed-end direction. However, the barbed end of the FH2–actin assembly is sterically blocked for ideal short-pitch and long-pitch addition to the post-site- and knob-site-bound monomers, respectively, and thus should be inhibited for growth (Figs 3a, left panel, and 5). Migration of the upper (blue) bridge element from actins 2/3

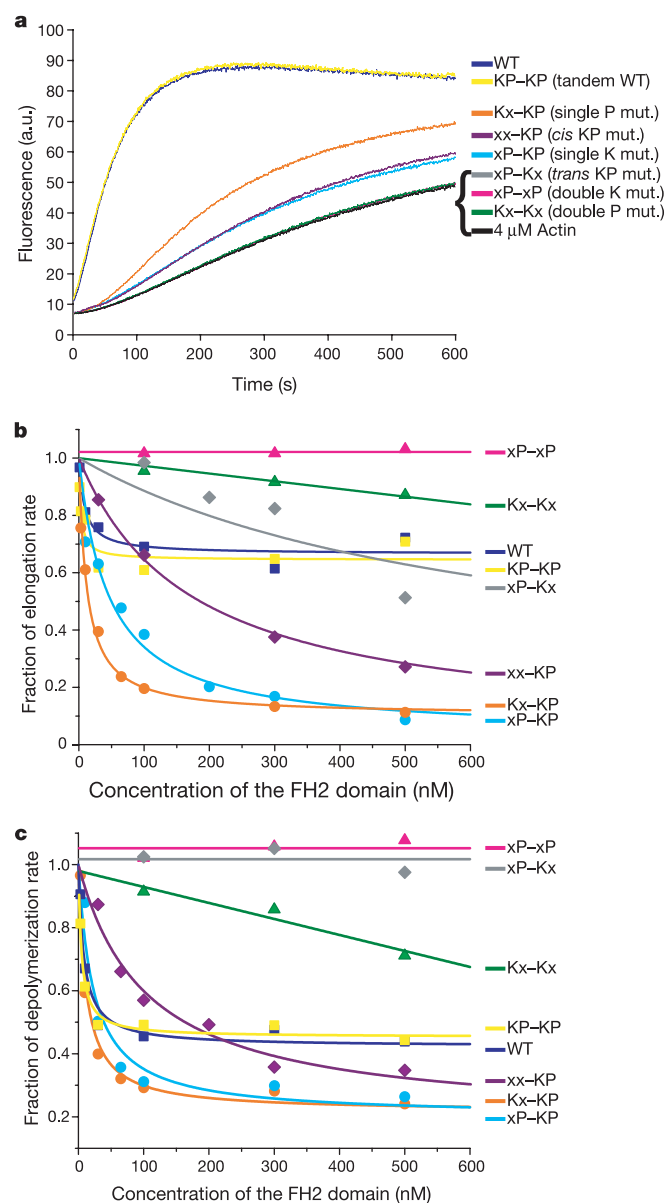


Figure 4 Actin regulation by linked FH2 domain mutants. **a**, Actin assembly assays with 30 nM mutant FH2 proteins (concentration defined according to bridge, for comparison to Supplementary Fig. S2c–e). a.u., arbitrary units; WT, wild type. **b**, Barbed-end elongation rate as a function of FH2 concentration (see Supplementary Fig. S5 for raw data). Elongation rates are normalized to the rate in the absence of FH2 domain. Curves were fit to a single site-binding isotherm to obtain K_D . **c**, Actin filament depolymerization rate in the presence of FH2 mutant proteins (see Supplementary Fig. S5 for raw data). The data were fitted as in **b**.

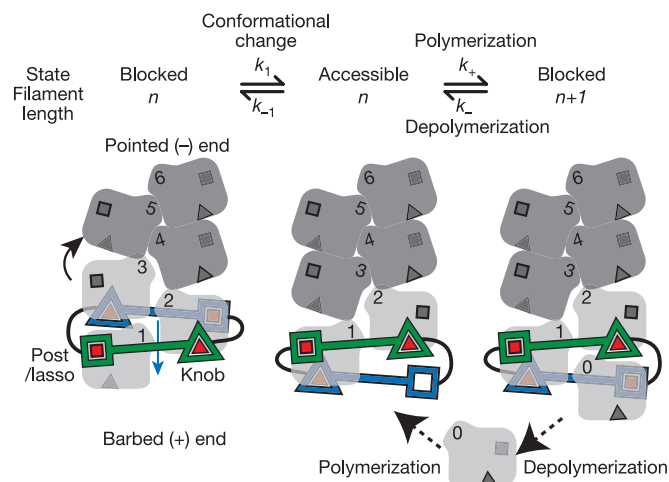


Figure 5 Model for FH2-mediated barbed end dynamics. Actin monomer is shown as a box with triangles and squares to indicate binding sites for the FH2 knob and post subdomains, respectively. FH2-bound sites are red, unbound sites are grey. The bridge element is shown as a bar with knob and post binding sites shaped to match the contact sites in actin. Bridges are blue and green, with black inter-bridge linkers. Actins are semi-transparent, to illustrate bridges and binding sites at the back of the structures. Actins held in the strained FH2-bound orientation are aligned with the page and light grey; actins in the filament orientation are tilted and dark grey. The configurational equilibrium between the closed and open states is determined by rate constants k_1 and k_{-1} . The intermolecular binding equilibrium between the two states is determined by rate constants k_+ and k_- and the concentration of actin monomer.

to actin 1 could relieve this inhibition by providing a free post site to recruit an incoming monomer (from left to centre panel in Fig. 5). This migration could occur with the FH2 dimeric ring intact; simple modelling demonstrates that the inter-bridge linker is long enough to pass around the sides of actins 1 and 3 easily without disrupting either set of lasso–post contacts within the formin (not shown). The rearranged bridge would then be attached to actin 1 only through its knob site. In concert with migration of the bridge, actin 3 would rearrange to a less strained, more filament-like orientation. Binding of new monomer to the free post site would then regenerate the original three-actin/two-bridge structure with the filament elongated by one monomer (actin 0). Repeated cycles of configurational exchange of the bridges and monomer binding would then lead to barbed-end growth in the presence of the FH2 cap, giving rise to the FH2–actin assembly attached to bulk filament, as described above.

In a general sense, our model proposes that the FH2 dimer exists in equilibrium between two states at the barbed end (Fig. 5). In the blocked state, all four sites in the two bridges are engaged and all three bound actins are in the strained conformation. In the accessible state, only three of the four binding sites are engaged, and the post site of the lower bridge is exposed in solution. In this case, only the two terminal actins are in the strained conformation. The transition from the blocked to the accessible state involves release of actin 3 from the upper bridge, allowing it to incorporate into the filament with more-optimal geometry. Thus, the equilibrium between these two configurations constitutes a balance between binding energy of the post site in the blocked state and relief of strain energy of actin 3 in the accessible state. During review of this manuscript, another study reported that the mDia1 FH2 domain could accelerate nucleotide hydrolysis during processive filament growth²¹. If this hydrolysis occurs at the terminal monomer(s), the actin nucleotide state could also contribute to the FH2 configurational equilibrium during elongation. This effect would not be relevant during filament depolymerization (because all monomers are ADP-bound), which also probably occurs processively based on kinetic considerations. The geometry of the bridge–actin complex appears to be precisely tuned to provide the balance of strain and binding energy needed to control actin monomer binding and dissociation from the barbed end, yet it also provides sufficient similarity to the natural filament to allow pointed-end attachment to the bulk filament. Notably, actin can efficiently add to the barbed end only in the accessible state (actin 0), because an exposed bridge post site is necessary for recruitment at this position. Actin can efficiently dissociate from the filament only in the blocked state, because in this configuration stabilizing long-pitch contacts between actins 0 and 2 are lost (see above), and actin 0 is bound to the filament through only one bridge site. The existence of both states and a dynamic equilibrium between them are thus necessary features of the Bni1p FH2 domain, allowing it to remain tightly bound to the filament barbed end, yet permitting processive filament growth and shrinkage in response to appropriate actin concentrations. In this model, transitions between the blocked and accessible states occur randomly (with relative probabilities determined by the energy difference between the states). However, when actin is above the critical concentration, once a monomer (actin 0) binds to the accessible state, the terminal bridge (blue) can no longer migrate back to the previous blocked configuration (that is to actins 2 and 3) due to the energetic penalty of destabilizing the new terminal monomer. Thus, the system behaves as a brownian ratchet²⁹, where fluctuations in FH2 configuration are captured and rectified through binding of a new actin monomer at the barbed end.

This model for processive capping can reasonably explain our biochemical data on the FH2 mutants. The xx–KP mutant could engage the two terminal monomers at the filament barbed end, holding them in the pseudo-short-pitch orientation through con-

tacts to the intact bridge element. Because the mutant lacks the second functional bridge element needed to recruit new monomer, it blocks filament growth. As a single bridge does not interact with actin 3, stronger actin 1–actin 3 long-pitch contacts could be made than with the full FH2 dimer, leading to slower filament depolymerization. The Kx–KP and xP–KP single mutants bind the barbed end more tightly than a single bridge element (Supplementary Table S1), indicating that an additional functional site in each probably engages the filament. The details of how these mutants are bound to the barbed end are not known; one possibility is that in order to maximize interactions of their actin-binding sites and stability of the filament, the Kx–KP and xP–KP mutants may differentially favour the accessible and blocked configurations in the FH2 equilibrium, respectively. Such biasing of the equilibrium, coupled with an inability to recruit new monomer through the post site (Kx–KP) or destabilize actin 1–actin 3 long pitch contacts through two knob sites (xP–KP) could explain the functional properties of these proteins. Direct structural analyses will be necessary to test these possibilities.

Variations in capping

Our model provides possible explanations for the functional variability among different formins^{11,14–18} and the effects of FH1-mediated recruitment of profilin–actin on FH2 function^{13,14,16,21} (Fig. 5).

At an FH2-bound barbed end, the actin-binding energetics derive from a combination of actin–actin and actin–FH2 contacts, with the latter probably dominant (see above). Thus, $k_+[actin]$ and k_- (the pseudo-first-order and first-order rate constants for monomer binding and dissociation, respectively) could vary among the FH2 family members. Additionally, elongation and depolymerization rates will be governed not only by the individual steps of monomer addition and release (k_+ and k_-), as in a free filament, but also by the rates and/or populations (depending on rate regime) involved in the configurational equilibrium of the FH2 domain (k_1 , k_{-1} , the first order rate constants for the forward and reverse conformational changes, and k_1/k_{-1}). These, too, could vary between different family members. Similarly, critical concentration is given by the product of both equilibria, not simply that of monomer addition and release ($[actin]_{crit} = (k_{-1}k_-)/(k_1k_+)$), and could vary as well. Together, these dictate that the thermodynamics and/or kinetics at the barbed end could in principle differ greatly between different formins due to sequence and structural variations. Finally, FH1-mediated recruitment of actin–profilin complexes could also change the monomer-binding rate by increasing the local actin concentration (and thus $k_+[actin]$) at the FH2-bound barbed end, providing rates even beyond that of free filaments, as recently observed²¹.

Discussion

Our model for processive capping of the actin filament is based on a structural unit, the FH2 actin bridge, that binds the barbed end and blocks monomer binding and release. In the FH2 dimer, the second bridge element provides functionality that can overcome the inhibitory properties of the first. A dynamic equilibrium of the bridges enables repeated binding or release of actin monomer from solution while protecting the filament from other capping proteins. Because individual steps in this model advance by only a single actin monomer, the twist of the filament should cause the FH2 domain to rotate during processive growth or shrinkage. This would seem to conflict with formin incorporation into large, membrane-anchored assemblies *in vivo*, and observations that immobilized filaments growing from immobilized FH2 domains do not supercoil¹⁹. We have considered several possible resolutions to this paradox. The first is that FH2 domains may be attached in these systems through linkages that allow rotation, either of membrane-bound components or of rotatable bonds in linkers. Alternatively, there could

be additional binding modes of the FH2 domain at the barbed end that are distinct from those in our crystal. Although this is possible in principle, numerous modes would probably be required to accommodate the 13 orientations in a half-turn of the long-pitch helix. Finally, the FH2 domain may function through a combination of non-dissociative and dissociative modes, with perhaps the latter dominating under tension in the filament. The structure and proteins described here provide tools to resolve this issue and the next set of mechanistic questions on formin function. □

Methods

Crystallization

Proteins were expressed and purified using standard methods, as described in the Supplementary Methods. The Bni1p FH2 domain in 10 mM HEPES, pH 7.0, 1 mM dithiothreitol (DTT), 100 mM KBr was mixed with TMR-actin in 2 mM Tris-HCl, pH 8.0, 0.2 mM ATP, 0.5 mM DTT, 0.1 mM CaCl₂ (buffer G) to yield 400–440 μM FH2 domain and 400 μM actin, and supplemented with 2 mM ATP and 5 mM CaCl₂. Crystals were grown at 20 °C by hanging-drop vapour diffusion by diluting 0.8–1 μl of the complex in an equal volume of 40 mM Tris-HCl, pH 7.6 and equilibrating against 600 μl of water. Crystals grew to 200 × 100 × 100 μm in 2 weeks, and were cryo-protected with 10 mM Tris-HCl, pH 7.6, 25 mM KBr, 35% ethylene glycol and flash-cooled in liquid propane. Crystals exhibit C2 symmetry with cell dimensions of $a = 232 \text{ Å}$, $b = 56 \text{ Å}$, $c = 101 \text{ Å}$, $\beta = 107.7^\circ$ and contain one molecule each of FH2 and actin in the asymmetric unit.

Data collection and structure determination

Diffraction data were collected at beamline 19-ID (SBC-CAT) at the Advanced Photon Source (Argonne National Laboratory) and processed with HKL2000³⁰. Phases were calculated from a three-wavelength anomalous dispersion (MAD) experiment at the selenium peak, inflection and high remote energies using an actin–FH2 crystal in which only the FH2 component was labelled with selenomethionine. Six of eight possible selenium sites were identified with the program CNS³¹ and refined with the program MLPHARE³². Electron density maps were calculated using MAD phases to 3.5 Å resolution and improved by density modification using DM³³. A native data set was collected to 3.05 Å resolution and used for model building (program O³⁴) and positional as well as grouped *B*-value refinement (CNS), resulting in an R_{work} of 28.9% and an R_{free} of 31.3%. The model contains residues 4–39 and 52–372 for actin as well as residues 1350–1760 for FH2, one ATP, and one Ca²⁺ ion. The regions in actin and the FH2 domain that interact with each other are best defined. Average *B*-values for actin are 91 Å² (subdomain 1), 126 Å² (subdomain 2), 97 Å² (subdomain 3) and 124 Å² (subdomain 4). For FH2, they are 121 Å² (lasso), 139 Å² (linker), 111 Å² (knob), 112 Å² (coiled coil) and 106 Å² (post). The model was verified using a composite simulated annealing omit map showing that all residues have contiguous electron density, except Ala 230 and Asn 252 of actin and Trp 1488 of the FH2 domain. Details of phase calculations and model refinement are provided in the Supplementary Methods.

Biochemical analyses

Assembly of 4 μM actin (5% pyrene labelled) at 25 °C was measured as reported³⁵. Elongation assays were performed as reported¹⁵. Briefly, FH2 domain in buffer F (buffer G plus 50 mM KCl, 1 mM MgCl₂, 1 mM EGTA, 10 mM imidazole pH 7.0) was mixed with pre-formed phalloidin actin filaments (~0.1 nM barbed ends), incubated for 2 min at 25 °C and added to Mg-actin (final 0.5 μM, 40% pyrene). Rates were obtained by linear regression of the first 100 s of pyrene fluorescence data. For depolymerization, 5 μM 70% pyrene actin was polymerized with 2 mM MgCl₂ and 50 mM KCl for 2 h, then diluted to a final concentration of 0.1 μM actin in buffer F in the presence of FH2 domains. Rates were determined by linear regression of fluorescence between 100 and 300 s.

Received 14 October; accepted 8 December 2004; doi:10.1038/nature03251.

Published online 5 January 2005.

1. Zigmond, S. H. Formin-induced nucleation of actin filaments. *Curr. Opin. Cell Biol.* **16**, 99–105 (2004).
2. Wallar, B. J. & Alberts, A. S. The formins: active scaffolds that remodel the cytoskeleton. *Trends Cell Biol.* **13**, 435–446 (2003).
3. Evangelista, M., Zigmond, S. & Boone, C. Formins: signaling effectors for assembly and polarization of actin filaments. *J. Cell Sci.* **116**, 2603–2611 (2003).
4. Evangelista, M. *et al.* Bni1p, a yeast formin linking cdc42p and the actin cytoskeleton during polarized morphogenesis. *Science* **276**, 118–122 (1997).
5. Evangelista, M., Pruyne, D., Amberg, D. C., Boone, C. & Bretscher, A. Formins direct Arp2/3-independent actin filament assembly to polarize cell growth in yeast. *Nature Cell Biol.* **4**, 260–269 (2002).
6. Sagot, I., Klee, S. K. & Pellman, D. Yeast formins regulate cell polarity by controlling the assembly of actin cables. *Nature Cell Biol.* **4**, 42–50 (2002).
7. Chang, F., Drubin, D. & Nurse, P. cdc12p, a protein required for cytokinesis in fission yeast, is a component of the cell division ring and interacts with profilin. *J. Cell Biol.* **137**, 169–182 (1997).
8. Chang, F. Movement of a cytokinesis factor cdc12p to the site of cell division. *Curr. Biol.* **9**, 849–852 (1999).

9. Watanabe, N., Kato, T., Fujita, A., Ishizaki, T. & Narumiya, S. Cooperation between mDia1 and ROCK in Rho-induced actin reorganization. *Nature Cell Biol.* **1**, 136–143 (1999).
10. Tominaga, T. *et al.* Diaphanous-related formins bridge Rho GTPase and Src tyrosine kinase signaling. *Mol. Cell* **5**, 13–25 (2000).
11. Pruyne, D. *et al.* Role of formins in actin assembly: nucleation and barbed-end association. *Science* **297**, 612–615 (2002).
12. Sagot, I., Rodal, A. A., Moseley, J., Goode, B. L. & Pellman, D. An actin nucleation mechanism mediated by Bni1 and profilin. *Nature Cell Biol.* **4**, 626–631 (2002).
13. Kovar, D. R., Kuhn, J. R., Tichy, A. L. & Pollard, T. D. The fission yeast cytokinesis formin Cdc12p is a barbed end actin filament capping protein gated by profilin. *J. Cell Biol.* **161**, 875–887 (2003).
14. Li, F. & Higgs, H. N. The mouse Formin mDia1 is a potent actin nucleation factor regulated by autoinhibition. *Curr. Biol.* **13**, 1335–1340 (2003).
15. Harris, E. S., Li, F. & Higgs, H. N. The mouse formin, FRLα, slows actin filament barbed end elongation, competes with capping protein, accelerates polymerization from monomers, and severs filaments. *J. Biol. Chem.* **279**, 20076–20087 (2004).
16. Pring, M., Evangelista, M., Boone, C., Yang, C. & Zigmond, S. H. Mechanism of formin-induced nucleation of actin filaments. *Biochemistry* **42**, 486–496 (2003).
17. Moseley, J. B. *et al.* A conserved mechanism for Bni1- and mDia1-induced actin assembly and dual regulation of Bni1 by Bud6 and profilin. *Mol. Biol. Cell* **15**, 896–907 (2004).
18. Zigmond, S. H. *et al.* Formin leaky cap allows elongation in the presence of tight capping proteins. *Curr. Biol.* **13**, 1820–1823 (2003).
19. Kovar, D. R. & Pollard, T. D. Insertional assembly of actin filament barbed ends in association with formins produces piconewton forces. *Proc. Natl Acad. Sci. USA* **101**, 14725–14730 (2004).
20. Higashida, C. *et al.* Actin polymerization-driven molecular movement of mDia1 in living cells. *Science* **303**, 2007–2010 (2004).
21. Romero, S. *et al.* Formin is a processive motor that requires profilin to accelerate actin assembly and associated ATP hydrolysis. *Cell* **119**, 419–429 (2004).
22. Xu, Y. *et al.* Crystal structures of a Formin Homology-2 domain reveal a tethered dimer architecture. *Cell* **116**, 711–723 (2004).
23. Shimada, A. *et al.* The core FH2 domain of diaphanous-related formins is an elongated actin binding protein that inhibits polymerization. *Mol. Cell* **13**, 511–522 (2004).
24. Graceffa, P. & Dominguez, R. Crystal structure of monomeric actin in the ATP state. Structural basis of nucleotide-dependent actin dynamics. *J. Biol. Chem.* **278**, 34172–34180 (2003).
25. Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. Atomic model of the actin filament. *Nature* **347**, 44–49 (1990).
26. Lorenz, M., Popp, D. & Holmes, K. C. Refinement of the F-actin model against X-ray fiber diffraction data by the use of a directed mutation algorithm. *J. Mol. Biol.* **234**, 826–836 (1993).
27. Galkin, V. E., VanLoock, M. S., Orlova, A. & Egelman, E. H. A new internal mode in F-actin helps explain the remarkable evolutionary conservation of actin's sequence and structure. *Curr. Biol.* **12**, 570–575 (2002).
28. Schmid, M. F., Sherman, P., Matsudaira, P. & Chiu, W. Structure of the acrosomal bundle. *Nature* **431**, 104–107 (2004).
29. Mogilner, A. & Oster, G. Polymer motors: pushing out the front and pulling up the back. *Curr. Biol.* **13**, R721–R733 (2003).
30. Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
31. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
32. Otwinowski, Z. in *Isomorphous Replacement and Anomalous Scattering* (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 80–86 (Science & Engineering Research Council, Cambridge, 1991).
33. Cowtan, K. & Main, P. Miscellaneous algorithms for density modification. *Acta Crystallogr. D* **54**, 487–493 (1998).
34. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
35. Marchand, J. B., Kaiser, D. A., Pollard, T. D. & Higgs, H. N. Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. *Nature Cell Biol.* **3**, 76–82 (2001).
36. DeLano, W. L. *The PyMOL User's Manual* (DeLano Scientific, San Carlos, California, 2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Hill and C. Brautigam for technical assistance; Y. M. Chook, M. Kikkawa, R. Ranganathan and D. Morgan for discussion and critical reading of the manuscript; S. Padrick for assistance with modelling FH2-mediated polymerization and depolymerization; and P. Mishra for help with light scattering experiments. This work was supported by grants from the NIH to M.K.R. T.O. was supported by the Human Frontier Science Program. Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the US Department of Energy, Office of Energy Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.K.R. (mrosen@biochem.swmed.edu). Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession code 1Y64.

The mass of the missing baryons in the X-ray forest of the warm-hot intergalactic medium

Fabrizio Nicastro¹, Smita Mathur², Martin Elvis¹, Jeremy Drake¹, Taotao Fang³, Antonella Fruscione¹, Yair Krongold^{1,4}, Herman Marshall⁵, Rik Williams² & Andreas Zezas¹

¹Harvard-Smithsonian Center for Astrophysics, 60 Garden St, Cambridge, Massachusetts 02138, USA

²Astronomy Department, The Ohio State University, Columbus, Ohio 43210, USA

³University of California-Berkeley, Berkeley, California, 94720 USA

⁴Istituto de Astronomia, UNAM, Mexico City, DF 04510, Mexico

⁵Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Recent cosmological measurements indicate that baryons comprise about four per cent of the total mass-energy density of the Universe^{1,2}, which is in accord with the predictions arising from studies of the production of the lightest elements³. It is also in agreement with the actual number of baryons detected at early times (redshifts $z > 2$)^{4,5}. Close to our own epoch ($z < 2$), however, the number of baryons detected add up to just over half (~ 55 per cent) of the number seen at $z > 2$ (refs 6–11), meaning that about ~ 45 per cent are ‘missing’. Here we report a determination of the mass-density of a previously undetected population of baryons, in the warm-hot phase of the intergalactic medium. We show that this mass density is consistent, within the uncertainties, with the mass density of the missing baryons.

Hydrodynamical simulations for the formation of structures in the Universe offer a possible solution to the puzzle of the ‘missing baryons’ (see refs 12–14). They predict that in the present epoch ($z \leq 1 - 2$), ~ 30 – 40% of the normal baryonic matter in the Universe lies in a tenuous Warm-Hot Intergalactic Medium (WHIM) at overdensities $\delta = n_b / \langle n_b \rangle \approx 5 - 100$ ($\langle n_b \rangle = 2 \times 10^{-7} (1+z)^3 (\Omega_b h^2 / 0.02) \text{ cm}^{-3}$ is the average density in the Universe), a factor 10^6 below the density of the interstellar medium in our Galaxy. The WHIM was shock-heated to temperatures of 10^5 – 10^7 K during the continuous process of structure formation^{12,13}. At these temperatures C, N, O and Ne (the most abundant metals in gas with solar or solar-like composition) are highly ionized, being mainly distributed between their He-like and H-like species. The strongest bound-bound transitions from these ions fall in the soft-X-ray band (for example, $\lambda(\text{O VII K}\alpha) = 21.602 \text{ \AA}$, and $\lambda(\text{O VIII K}\alpha) = 18.97 \text{ \AA}$), and so a ‘forest’ of weak (that is, $N < 10^{16} \text{ cm}^{-2}$, or equivalent width, $W < 18 - 30 \text{ m\AA}$, depending on the particular ion) absorption lines is expected to be imprinted in the X-ray spectra of all background sources at redshifts between zero and ~ 1 – 2 by these WHIM filaments (see refs 14, 15), as direct analogues of the ‘Lyman- α forest’ due to cool neutral hydrogen detected in the optical spectra of quasars. The Lyman- α forest accounts for virtually all the baryons at $z > 2$ (refs 4, 5), but only for about 30% of them at lower z (ref. 9; Table 1).

Owing to our special embedded location, metal absorption lines from warm-hot gas at $z = 0$ (either Local Group WHIM¹⁶ or an extended Galactic halo^{17,18}) have been detected in X-rays^{19–22} and in the ultraviolet^{16,17}. However, even if all this matter were to be identified with Local Group WHIM, our preferential location would not allow us to infer the WHIM contribution to the cosmological baryon density in the local Universe. To measure Ω_b (WHIM) requires observation along a random, unprivileged line of sight. Owing to the intrinsic steepness of the number of absorbers as a function of ion column density for the WHIM metal absorption lines, they are rare at relatively large columns (see ref. 14, for example). Only a single O VII K α system with $N_{\text{O VII}} \geq 10^{16} \text{ cm}^{-2}$,

where N is the column density (in cm^{-2}) is expected along a random line of sight, up to $z = 0.3$, with a probability of $\sim 60\%$, whereas eight systems with $N_{\text{O VII}} \geq 10^{15} \text{ cm}^{-2}$ are expected. So detecting weaker lines is expected to be more productive than probing higher redshifts.

In practice, a very large number of counts per resolution element is needed to detect these lines in current spectra from the Chandra low energy transmission grating (LETG) or the XMM-Newton reflection grating spectrometer (RGS), both of which have a resolution of $50 \text{ m\AA}$ ($R \approx 400$ at $\sim 20 \text{ \AA}$). To probe the more common WHIM columns of $N_{\text{O VII}} \approx 10^{15} \text{ cm}^{-2}$, which have equivalent widths of only $\sim 3 \text{ m\AA}$, about 2,500 counts per resolution element are needed for a 3σ detection of a non-saturated line. The brightest Seyfert galaxies and blazars have typical soft X-ray (0.5–2 keV) fluxes of the order of $4 \times 10^{-11} \text{ erg s}^{-1} \text{ cm}^{-2}$ (2 mCrab), which would take a 2.5-Ms Chandra-LETG exposure to reach this signal-to-noise ratio (S/N). Furthermore, all these objects are nearby ($z \leq 0.05$), and so sample path-lengths of only a few hundred megaparsecs. Indeed, no secure X-ray detection of higher-redshift, intervening WHIM absorption systems has been claimed so far²³. One solution is to observe background sources in unusually bright states.

On 26–27 October 2002 and 1–2 July 2003 we triggered, under our Chandra-AO4 programme, two 100-ks LETG observations of the blazar Markarian 421. These two observations caught the source at historical maxima of 60 and 40 mCrab in the 0.5–2-keV band, allowing us to collect the best S/N of any high-resolution spectrum taken with Chandra to date, including Galactic X-ray binaries. We detected 24 metal absorption lines in the LETG spectrum of Mkn 421 (Fig. 1), with conservative significances $> 2\sigma$ (for additional details, please see the Supplementary Methods). We identify nine of these 24 absorption lines as belonging to two different intervening absorption systems, at $z = (0.011 \pm 0.001)$ and $z = (0.027 \pm 0.001)$ (Table 2). The remaining lines are identified as belonging to either the interstellar medium (ISM) of our Galaxy or the Local Group WHIM filament, and will not be discussed further here.

An intervening H I Ly α system at $cz = (3,046 \pm 2) \text{ km s}^{-1}$ ($z = 0.010160 \pm 0.000007$) was discovered in the Hubble Space Telescope (HST) Goddard High Resolution Spectrograph (GHRS) spectrum of Mkn 421, and reported in ref. 24. The redshift of the H I Ly α line is consistent with that of the $z = 0.011$ X-ray absorption system within the large systematic X-ray uncertainties ($\Delta z = 0.001$), but its full-width at half-maximum (FWHM) implies

Table 1 Census of baryons in the high-and low-redshift Universe

Inferred from	Ω_b (%) for $h_{70} = 1$	References
BBN + D/H*	(4.4 ± 0.4)	3
CMB anisotropy	(4.6 ± 0.2)	1, 2
Observed at $z > 2$ in†		
Lyman- α forest	> 3.5	4, 5
Observed at $z < 2$ in‡		
Stars	(0.26 ± 0.08)	6
H I + He I + H ₂	(0.080 ± 0.016)	6
X-ray gas in clusters	(0.21 ± 0.06)	6
Lyman- α forest	(1.34 ± 0.23)	9
Warm + warm-hot O VI	$(0.6^{+0.4}_{-0.3})$	10, 11
Total (at $z < 2$)§	$(2.5^{+0.5}_{-0.4})$	
Missing baryons (at $z < 2$)§	$(2.1^{+0.5}_{-0.4})$	

*The cosmological baryon density inferred by Big Bang nucleosynthesis (BBN) compared with observations of light element ratios, and observations of cosmic microwave background (CMB) anisotropies.

†The cosmological baryon density found in the Lyman- α forest at $z > 2$. Virtually all baryons are found in this cool component at high redshift.

‡The cosmological baryon density observed at $z < 2$ in virialized structures (stars, neutral H and He and molecular H and X-ray-emitting hot gas in clusters of galaxies and big groups of $M > 4 \times 10^{13} M_\odot$) and not-yet-virialized structures (cool gas in the local Lyman- α forest and warm-hot gas in the O VI WHIM).

§The total observed cosmological baryon density at $z < 2$ and its difference from the cosmological baryon density inferred by observations of CMB anisotropies, that is, the so-called ‘missing’ baryons.

a 3σ upper limit on the temperature of the H I absorber of $T_{\text{H I}}^{\text{max}} < 1.2 \times 10^5 \text{ K}$, inconsistent with the range of temperatures estimated for the X-ray absorber (see Fig. 2 legend). If the systems are related, a multiphase WHIM is required. No H I Ly α lines are visible at $z = 0.011$ and $z = 0.027$.

We also searched the Far Ultraviolet Spectrometer Explorer (FUSE) spectrum of Mkn 421 for O VI absorption at or close to

the redshifts of our two X-ray absorbers. No O VI lines are visible at $z = 0.010160$, $z = 0.011$ or $z = 0.027$. The conservative significances of the X-ray absorbers at $z = 0.011$ and $z = 0.027$ are 3.5σ and 4.9σ respectively, while their direct integrated significances are 5.8σ and 8.9σ (Table 2).

For both the $z > 0$ X-ray systems we compared the measurements (X-rays) or 3σ upper limits (ultraviolet, far-ultraviolet and

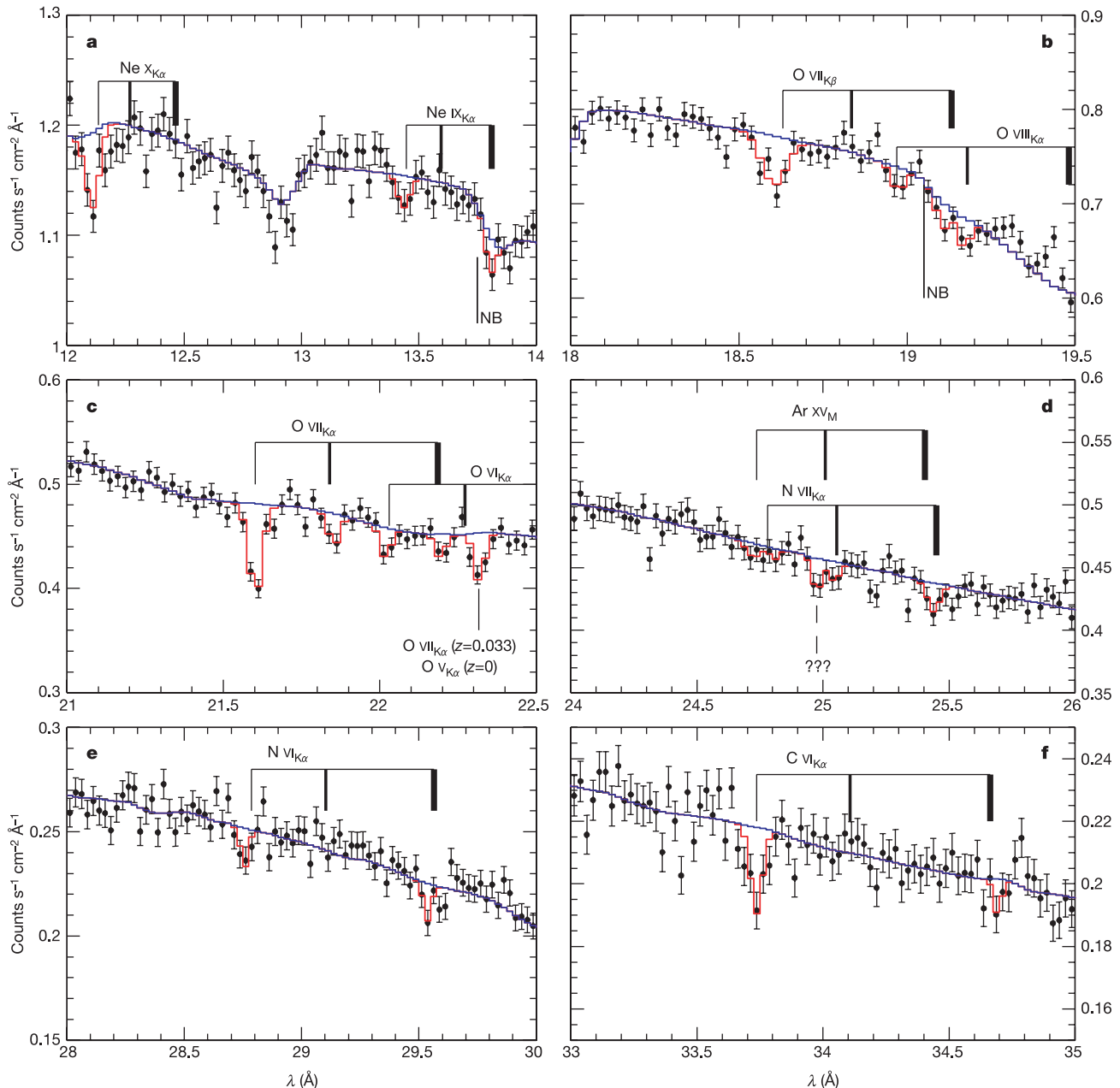


Figure 1 The WHIM absorption in the Chandra LETG spectrum of Mkn 421. Six portions of the LETG spectrum of Mkn 421 along with its best-fitting continuum (solid blue line) plus narrow absorption model (solid red line), centred around the rest wavelengths of the Ne IX–Ne X K_{α} (a), O VIII K_{α} (b), O VII K_{α} (c), N VII K_{α} (d), N VI K_{α} (e) and C VI K_{α} (f) transitions. This spectrum contains a total of $\sim 5,000$ counts per resolution element in the continuum at $21 \text{ \AA}$, enough to detect O VII columns of $N_{\text{O VII}} \geq 8 \times 10^{14} \text{ cm}^{-2}$ at a significance level $\geq 3\sigma$. For each labelled ion in the six panels, the three vertical lines from left to right are the rest frame wavelengths (thin line) and the expected wavelengths at $z = 0.011$ ($cz = 3,300 \text{ km s}^{-1}$, medium thickness line) and $z = 0.027$ ($cz = 8,090 \text{ km s}^{-1}$, thick line). Mkn 421 was also observed with the HST GHRS on 1 February 1995. We retrieved

this GHRS spectrum of Mkn 421 from the public HST archive, and re-analysed the data. The 3σ H I column upper limit of putative H I Ly α at the average redshifts of the two X-ray systems are $N_{\text{H I}} < 4.7 \times 10^{12} \text{ cm}^{-2}$ and $N_{\text{H I}} < 8.5 \times 10^{12} \text{ cm}^{-2}$ (assuming a temperature of $\log[T(\text{K})] = 6.1$ for both systems). Following our ‘target of opportunity observation’ request Mkn 421 was observed by FUSE on 19–21 January 2003 with a total exposure time of 62 ks. From this spectrum we derived 3σ O VI column upper limits of putative O VI $_{2s-2p}$ at the redshift of the H I Ly α absorber and at the average redshifts of the two X-ray systems, of $N_{\text{O VI}}(z = 0.010160) < 1.4 \times 10^{13} \text{ cm}^{-2}$, $N_{\text{O VI}}(z = 0.011) < 1.6 \times 10^{13} \text{ cm}^{-2}$ and $N_{\text{O VI}}(z = 0.027) < 1.4 \times 10^{13} \text{ cm}^{-2}$ respectively. Error bars, $\pm 1\sigma$

Table 2 The two WHIM filaments along the line of sight to Mkn 421

Redshift*	>2 σ ion detections†	Probability of exceeding F ‡	Conservative significance‡	Sum of lines significance§
0.011 ± 0.0011	O VII, N VII	5.9×10^{-4}	3.5σ	5.8σ
0.027 ± 0.0011	O VII, N VII, N VI	1.3×10^{-6}	4.9σ	8.9σ

*The redshift of the two X-ray absorbers and its error, which include a systematic 20-mÅ uncertainty in the LETG wavelength calibration.

†The ions detected at a significance >2 σ . C VI is detected in the $z = 0.027$ system, at a significance of 1.8 σ . O VIII and Ne IX are detected in the $z = 0.011$ and $z = 0.027$ systems respectively, at significance larger than 3 σ , but both detections are considered here as 3 σ upper limits owing to the proximity of these lines with known instrumental artefacts.

‡The total conservative significance of the two systems was estimated by comparing the best-fit χ^2 of simpler models that do not include the metal absorption lines from these systems with that of more complex models that include them. The F -tests were run over the following two portions of the spectrum, containing a similar number (~100) of resolution elements: $\Delta\lambda(z = 0.011) = [21, 26] \text{ \AA}$ and $\Delta\lambda(z = 0.027) = [21.5, 23] \cup [24.5, 26] \cup [27.5, 30] \cup [34, 35] \text{ \AA}$, where the union of the above intervals makes up the portion of the spectrum over which the test relative to the $z = 0.027$ filament have been run. O VII K_{α} and N VII K_{α} lines were included in the complex model for the $z = 0.011$ system, and O VII K_{α} , N VII K_{α} , N VI K_{α} and C VI K_{α} were used for the $z = 0.027$ system. The width of the lines was frozen to the instrument resolution, while their relative positions and normalizations were fixed to their expected relative position based on the proposed identifications, and their measured relative normalization. Only the redshift of the system and the overall normalization were left free to vary in the fits.

§The sum of the significances of the lines listed in column 2 (each computed as the ratio between measured equivalent width W and its 3 σ error: see Supplementary Methods).

X-rays) of the equivalent widths and equivalent width ratios with ion relative abundances and ratios from hybrid 'collisional ionization plus photoionization' models (see Supplementary Methods for details), to derive the physical properties of the gas (for example, see ref. 19).

For both systems we find solutions that define physical, geometrical and astrophysical parameters consistent with theoretical predictions for the WHIM (see Fig. 2 legend for details). We therefore identify these two intervening absorbers with two tenuous WHIM filaments absorbing the ultraviolet and X-ray radiation of Mkn 421 along our line of sight (see the online Supplementary Discussion for the arguments used to rule out the identification of the absorbers with reprocessed material flowing out from the blazar environment or its host galaxy).

Previous works have attempted to estimate the baryon content of the O VII and O VIII WHIM on the basis of either X-ray absorption line upper limits²⁵ or single-line detections²⁶ (none of which were confirmed by subsequent observations^{19,22}). These estimates all suffer the large uncertainties due to the poor ionization corrections used and the absence of a metallicity estimate. From the two WHIM detections presented here we can, for the first time, estimate the number density of O VII WHIM filaments with $N_{\text{O VII}}^{\text{thresh}} \geq 7 \times 10^{14} \text{ cm}^{-2}$ (the weakest O VII column that we detect, in the $z = 0.027$ filament) in the local Universe. We find $dP/dz = 67_{-43}^{+88}$ (where dP/dz is the number of filaments per unit redshift), which is roughly four times the number density of the ~17.5 times weaker O VI absorbers²⁸, and is consistent within the 1 σ errors²⁹ with hydrodynamical simulation predictions, $(dP/dz)_{\text{expected}} = 30.1$ (Fig. 2)^{13,14}.

More importantly, the accurate ionization correction and metallicity estimates enabled by O VI–O VIII, N VI–N VII and H I ratios allow us to derive a first estimate of the mean cosmological baryon density in X-ray filaments. Following ref. 27 (and approximating their formula (1) to the low-redshift regime), we estimate the baryon mass density of the WHIM, in units of critical density ρ_c , as:

$$\Omega_b = \left(\frac{1}{\rho_c} \right) \left(\frac{\mu m_p \sum_i N_{\text{H}}^i}{d_{\text{Mkn 421}}} \right)$$

where μ is the average molecular mass relative to the mass of the proton (taken to be 1.3), N_{H}^i is the equivalent H column density of the WHIM filament i and $d_{\text{Mkn 421}} = (zc/H_0) = 128 h_{70}^{-1} \text{ Mpc}$ is the distance to Mkn 421. Adopting a central temperature of $\log[T(\text{K})] = 6.1$ for both systems gives $\Omega_b^{\text{WHIM}} = (2.7_{-1.9}^{+3.8} \%) \times 10^{-10/\text{H}-1}$ (here we included both the measured uncertainties on

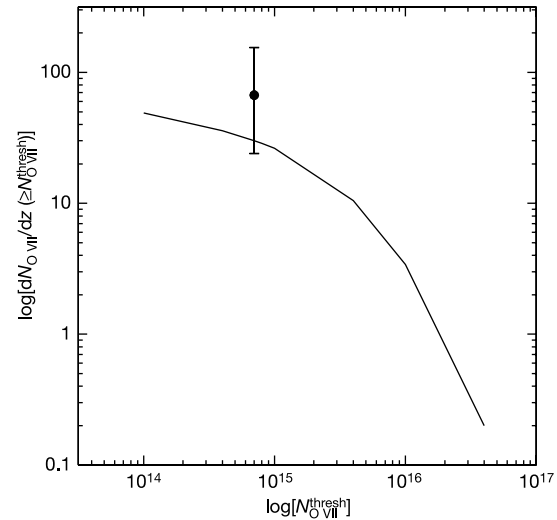


Figure 2 The 'missing' baryons have been found in the WHIM. Predicted curves $dN_{\text{O VII}}/dz (\geq N_{\text{O VII}}^{\text{thresh}})$ versus $N_{\text{O VII}}^{\text{thresh}}$ (ref. 14). The data point $[dN_{\text{O VII}}/dz (N_{\text{O VII}} \geq 7 \times 10^{14})]_{\text{observed}} = 67_{-43}^{+88}$ is also shown with $\pm 1\sigma$ errors, and is consistent with the corresponding theoretical expectation value $[dN_{\text{O VII}}/dz (N_{\text{O VII}} \geq 7 \times 10^{14})]_{\text{expected}} = 30.1$. All physical, geometrical and astrophysical parameters of the two observed intervening systems are also fully consistent with WHIM expectations. For the system at $z = 0.011$ we find common N_{H} solutions in a relatively broad range of temperatures: $\log[T(\text{K})] = (5.8-6.4)$, and constrain the oxygen metallicity to $[O/H] > -1.47$ (note that, to find these solutions, we use the 3 σ upper limit on the column of the putative H I Ly α at $z = 0.011$, and not the measured column of the H I Ly α at $z = 0.010160$, since the temperature implied by the width of this line is inconsistent with the range of temperatures that define the X-ray solution). For the system at $z = 0.027$ we find that solutions exist only for a relatively narrow ($\Delta T \approx 3 \times 10^5 \text{ K}$) range of temperatures around $T = 1.4 \times 10^6 \text{ K}$, and only for rather high N/O ratios compared to solar: $[N/O] \geq 0.65$, and again moderately high metallicity $[O/H] \geq -1.32$. The equivalent H column densities of the two intervening systems are $N_{\text{H}}(z = 0.011) = [(1.7 \pm 0.7) T_6 - (0.1 \pm 0.6)] \times 10^{19} \times 10^{-10/\text{H}-1} \text{ cm}^{-2}$ and $N_{\text{H}}(z = 0.027) = [(1.67 \pm 0.76) T_6 - (0.71 \pm 1.15)] \times 10^{19} \times 10^{-10/\text{H}-1} 10^{-[N/O]_{0.65}} \text{ cm}^{-2}$. These, in turn, imply thicknesses of the absorbers along the line of sight of $D(z = 0.011) = [(0.55 \pm 0.23) T_6 - (0.03 \pm 0.19)] (n_{-5b})^{-1} 10^{-10/\text{H}-1} \text{ Mpc}$ and $D(z = 0.027) = [(0.53 \pm 0.24) T_6 + (0.23 \pm 0.37)] (n_{-5b})^{-1} 10^{-10/\text{H}-1} 10^{-[N/O]_{0.65}} \text{ Mpc}$ (where n_{-5b} is the baryon volume density in units of 10^{-5} cm^{-3}).

N_{H} and the small number statistical errors for a poissonian distribution²⁸, but did not attempt to estimate the size of the uncertainty due to cosmic variance, because we have only observed a single line of sight). This is consistent with the expected $(\Omega_b^{\text{WHIM}})_{\text{predicted}} = 1.6\%-1.8\%$ (refs 13, 14), and with the total amount of 'missing' baryons $(\Omega_b)_{\text{missing}} = (2.1_{-0.4}^{+0.5} \%)$ (Table 1). □

Received 15 July; accepted 30 November 2004; doi:10.1038/nature03245.

- Bennett, C. L. *et al.* First-year Wilkinson microwave anisotropy probe (WMAP) observations: preliminary maps and basic results. *Astrophys. J. Suppl.* **148**, 1–27 (2003).
- Spergel, D. N. *et al.* First-year Wilkinson microwave anisotropy probe (WMAP) observations: determination of cosmological parameters. *Astrophys. J. Suppl.* **148**, 175–194 (2003).
- Kirkman, D., Tytler, D., Suzuki, M., O'Meara, J. M. & Lubin, D. The cosmological baryon density from the deuterium-to-hydrogen ratio in QSO absorption systems: D/H toward Q1243+3047. *Astrophys. J. Suppl.* **149**, 1–28 (2003).
- Rauch, M. The Lyman Alpha forest in the spectra of QSOs. *Annu. Rev. Astron. Astrophys.* **36**, 267–316 (1998).
- Weinberg, D. H., Miralda-Escude, J., Hernquist, L. & Katz, N. A lower bound on the cosmic baryon density. *Astrophys. J.* **490**, 564–570 (1997).
- Fukugita, M. Cosmic matter distribution: cosmic baryon budget revisited. In *Dark Matter in Galaxies* (Proc. IAU Symp. 220, in the press); preprint at (<http://www.arXiv.org/astro-ph/0312517>) (2003).
- Shull, J. M. In *The IGM/Galaxy Connection* (eds Rosenberg, J. L. & Putman, M. E.) Vol. 281, 1–10 (Kluwer Academic, Dordrecht, 2003).
- Stoeck, J. T., Shull, J. M. & Penton, S. V. The baryon content of the local intergalactic medium. *Proc. 2004 STScl May Symp. on Planets to Cosmology: Essential Science in Hubble's Final Years* (Cambridge Univ. Press, in the press); preprint at (<http://www.arXiv.org/astro-ph/0407352>) (2004).

9. Penton, S. V., Stocke, J. T. & Shull, J. M. The local Ly α forest. IV. Space Telescope Imaging Spectrograph G140M spectra and results on the distribution and baryon content of HI absorbers. *Astrophys. J. Suppl.* **152**, 29–62 (2004).
10. Tripp, T. M., Savage, B. D. & Jenkins, E. B. Intervening O VI quasar absorption systems at low redshift: a significant baryon reservoir. *Astrophys. J.* **534**, L1–L5 (2000).
11. Savage, B. D., Sembach, K. R., Tripp, T. M. & Richter, P. Far ultraviolet spectroscopic explorer and Space Telescope Imaging Spectrograph observations of intervening O VI absorption line systems in the spectrum of PG 0953+415. *Astrophys. J.* **564**, 631–649 (2002).
12. Cen, R. & Ostriker, J. P. Where are the baryons? *Astrophys. J.* **514**, 1–6 (1999).
13. Davé, R. *et al.* Baryons in the warm-hot intergalactic medium. *Astrophys. J.* **552**, 473–483 (2001).
14. Fang, T., Bryan, G. L. & Canizares, C. R. Simulating the X-ray forest. *Astrophys. J.* **564**, 604–623 (2002).
15. Perna, R. & Loeb, A. X-ray absorption by the hot intergalactic medium. *Astrophys. J.* **503**, L135–L138 (1998).
16. Nicastro, F. *et al.* The far-ultraviolet signature of the ‘missing’ baryons in the Local Group of galaxies. *Nature* **421**, 719–721 (2003).
17. Sembach, K. R. *et al.* Highly ionized high-velocity gas in the vicinity of the galaxy. *Astrophys. J. Suppl.* **146**, 165–208 (2003).
18. Collins, J. A., Shull, J. M. & Giroux, M. L. Highly ionized high-velocity clouds toward PKS 2155–304 and Markarian 509. *Astrophys. J.* **605**, 216–229 (2004).
19. Nicastro, F. *et al.* Chandra discovery of a tree in the X-ray forest toward PKS 2155–304: The local filament? *Astrophys. J.* **573**, 157–167 (2002).
20. Fang, T., Sembach, K. R. & Canizares, C. R. Chandra detection of local O VII He α absorption along the sight line toward 3C 273. *Astrophys. J.* **586**, L49–L52 (2003).
21. Rasmussen, A., Kahn, S. M. & Paerels, F. *X-ray IGM in the Local Group, The IGM/Galaxy Connection* (eds Rosenberg, J. L. & Putman, M. E.) 109–115 (Kluwer Academic, Dordrecht, 2003).
22. Cagnoni, I., Nicastro, F., Maraschi, L., Treves, A. & Tavecchio, F. A view of PKS 2155–304 with XMM-Newton reflection grating spectrometers. *Astrophys. J.* **603**, 449–455 (2004).
23. Nicastro, F. in *Chandra and FUSE View of the WHIM: The Local Group and Beyond, Maps of the Cosmos (14–17 July 2003, Sydney, Australia)* (eds Colless, M. & Staveley-Smith, L.) 170–177 (International Astronomical Union Symp. 216, IAU, 2003).
24. Shull, J. M., Stocke, J. T. & Penton, S. Intergalactic hydrogen clouds at low redshift: connections to voids and dwarf galaxies. *Astron. J.* **111**, 72–77 (1996).
25. Mathur, S., Weinberg, D. H. & Chen, X. Tracing the warm-hot intergalactic medium at low redshift: X-ray forest observations toward H1821+643. *Astrophys. J.* **582**, 82–94 (2003).
26. Fang, T., Marshall, H. L., Lee, J. C., Davis, D. S. & Canizares, C. R. Chandra detection of O VIII Ly α absorption from an overdense region in the intergalactic medium. *Astrophys. J.* **572**, L127–L130 (2002).
27. Savage, B. D., Sembach, K. R., Tripp, T. M. & Richter, P. Far Ultraviolet Spectroscopic Explorer and Space Telescope Imaging Spectrograph observations of intervening O VI absorption line systems in the spectrum of PG 0953+415. *Astrophys. J.* **564**, 631–649 (2002).
28. Gehrels, N. Confidence limits for small numbers of events in astrophysical data. *Astrophys. J.* **303**, 336–346 (1986).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work has been supported in part by NASA-Chandra and NASA-LTSA grants.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.N. (fnicastro@cfa.harvard.edu).

Creation of visible artificial optical emissions in the aurora by high-power radio waves

Todd R. Pedersen¹ & Elizabeth A. Gerken²

¹Space Vehicles Directorate, Air Force Research Laboratory, Hanscom Air Force Base, Massachusetts 01731, USA

²Department of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14853, USA

Generation of artificial light in the sky by means of high-power radio waves interacting with the ionospheric plasma has been envisaged since the early days of radio exploration of the upper atmosphere, with proposed applications ranging from regional night-time street lighting to atmospheric measurements¹. Weak optical emissions have been produced for decades in such

ionospheric ‘heating’ experiments, where they serve as key indicators of electron acceleration, thermal heating, and other effects of incompletely understood wave–particle interactions in the plasma under conditions difficult to replicate in the laboratory². The extremely low intensities produced previously have, however, required sensitive instrumentation for detection, preventing applications beyond scientific research. Here we report observations of radio-induced optical emissions bright enough to be seen by the naked eye, and produced not in the quiet mid-latitude ionosphere, but in the midst of a pulsating natural aurora. This may open the door to visual applications of ionospheric heating technology or provide a way to probe the dynamics of the natural aurora and magnetosphere.

The most readily observed emissions produced in ionospheric heating are the ‘forbidden’ red and green lines from atomic oxygen at 630.0 and 557.7 nm, both common components of the natural aurora and airglow³. In almost all past experiments, artificial emissions have been produced by the interaction of radio waves with the ionospheric F region, the long-lived primary ionospheric layer composed of atomic ions at an altitude of several hundred kilometres⁴. Only rarely have optical effects been reported from the ionospheric E region⁵, an ephemeral layer created from occasional meteoric ions or continuous solar illumination or auroral precipitation near an altitude of 100 km (ref. 4), where increased collisions with neutral molecules alter the behaviour of the plasma, and the proximity to the transmitter provides a large inverse-square increase in power density⁵. Emission intensities achieved previously have typically ranged up to several hundred Rayleighs (1 R = 10⁶ photons cm^{−2} s^{−1} integrated along a line of sight) for the more easily excited red line and tens of Rayleighs for the higher-energy green line. In all cases, intensities have remained far below

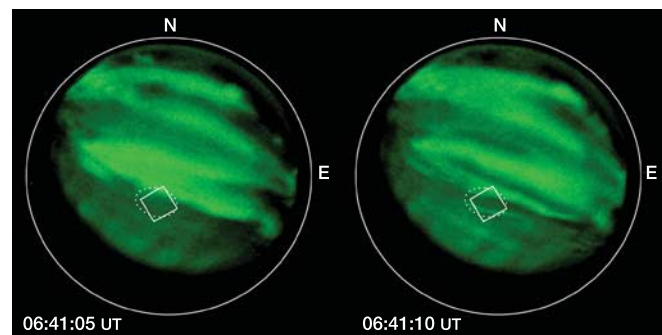


Figure 1 A pair of all-sky images taken 5 s apart showing the dynamic natural background aurora pulsating in long bands over most of the sky. These images were taken at 557.7 nm during the minute of 06:41 UT on 10 March 2004, when the HAARP transmitter was operating in O-mode polarization at a frequency of 5.95 MHz, providing 95 MW effective radiated power (ERP) directed at the magnetic zenith (azimuth 204°, elevation 75°). The half-power contour of the beam, represented by the dotted oval, enclosed a region approximately 15° by 21°, or 26 km by 37 km at an altitude of 100 km. The solid circle indicates the horizon, and the field of view of the higher-resolution imager is indicated by a white square. The pulsating auroral bands shown here were observed near 20:00 magnetic local time as part of the declining phase of an auroral event that had already persisted for several hours and reached more than 400 nT at its peak. The K_p planetary magnetic activity index was 6 for the three-hour periods before and after 6 UT. Auroral precipitation during this hour produced E layers ranging from ~4–6 MHz in critical frequency: $(2.0\text{--}4.5) \times 10^{11}$ electrons m^{−3}. The E layer at this time was centred at an altitude of about 110 km with a peak frequency of ~6.0 MHz and a half-thickness of 10–15 km. The nominal reflection altitude at the 5.95-MHz transmitter frequency was 109 km. Note that the time constant for changes in the E-region density is about 10 s (ref. 9), compared to the several minutes required to collect an ionogram, so nominal parameters cannot be expected accurately to represent instantaneous conditions in this highly dynamic environment.

the threshold for detection by the human eye, which is given as 20 kR at 630 nm (ref. 2) towards the red end of the visible wavelength range, and 1 kR for 558 nm (ref. 6) near the peak sensitivity of the eye.

We recently produced dramatically stronger artificial optical emissions bright enough to be visible to the naked eye in an experiment targeting the ionospheric E layer created by the natural aurora. The experiment was conducted on 10 March 2004, between ~6–7 UT, using the 960-kW transmitter array at the High Frequency Active Auroral Research Program (HAARP) facility near Gakona, Alaska (62.4° N, 145.15° W). The HAARP transmitter was run in a 15-s cycle alternating between 7.5 s of full power and 7.5 s off. Four filtered optical imaging systems ranging from all-sky to telescopic were operated in synchronization with the transmitter on and off intervals. Background conditions during the experiment period were characterized by aurora pulsating with apparent periods of ~10 s in longitudinal bands running in the magnetic east–west direction over most of the sky, including the region within the transmitter beam (Fig. 1). The auroral precipitation created a blanketing E layer near an altitude of 100 km with critical frequencies ranging from 4–6 MHz.

For a period of approximately 10 min between about 06:40 and 06:50, a number of small speckles of enhanced green emission were observed with the HAARP telescope wide-field camera, which provided high-resolution images of the region within the transmitter beam near magnetic zenith (Fig. 2). The speckles were present only during the image frames when the transmitter was on and were absent from exposures taken during the off periods. There is

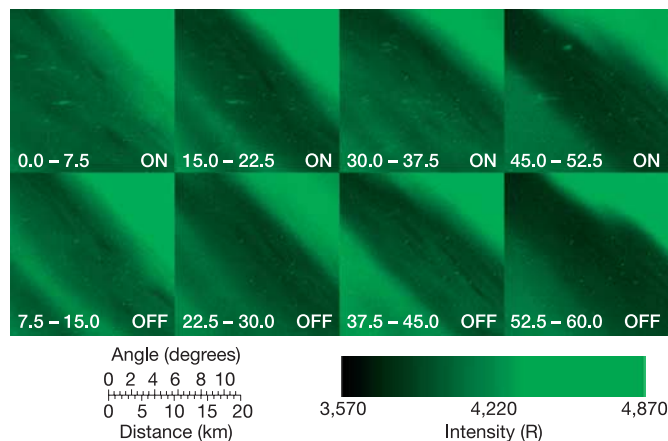


Figure 2 Images from the HAARP telescope wide field camera showing speckle-like artificial optical emissions superimposed on the background aurora only during frames when the transmitter was on. These data were taken at 557.7 nm during the minute of 06:41 UT on 10 March 2004. Frames taken during the 7.5 s the HAARP transmitter was turned on are in the top row with 'off' images along the bottom row. The camera was pointed at the magnetic zenith and covered a field of view of about 12° across, corresponding to a distance of about 20 km at an altitude of 100 km of the auroral E layer. Each image was integrated for just over 7 s so that with the data readout time the system would start each exposure at the beginning of each 7.5-s on or off period, indicated at lower left. True zenith is towards upper right and west is towards lower right. The speckles are visible in the upper row as bright spots much bigger than the stars and noise pixels. The larger speckles are approximately 30 pixels, or one degree, across, corresponding to a true size of about 1.75 km at the assumed 100-km altitude of the emission. The elongated shape and apparent displacement of many of the speckles are consistent with motion of the features of the order of several hundred metres per second of exposure time. Some frames, including those shown here, show small darker bands in the background suggestive of black aurora¹⁰, but as speckles are present in other frames lacking any apparent black aurora there does not appear to be any direct correlation.

evidence of dynamic pulsations in the background aurora within this narrower field of view as well, such as the auroral bands that appear and disappear in the lower left corner of the images. The largest speckles are approximately one degree across.

Within a given frame the speckles appear to be randomly distributed, but upon closer examination of successive 'on' frames, some of the speckles often appear to be correlated with but displaced from those seen in the previous 'on' period. This suggests that some speckle features are in motion but may still retain coherence across multiple on–off cycles of the transmitter.

Calibrated average intensities for the background aurora within the transmitter beam were obtained from another imager, which made measurements at several different wavelengths once each minute (Fig. 3). In spite of the rapid pulsations in narrow bands and on 10-s timescales, the average auroral brightness at 557.7 nm remained near 4 kR, with an increase to ~5 kR near the time the speckles were observed. This intensity calibration, applied to the high-resolution data in Fig. 2, indicates that the brightest speckles were approximately 4 kR in total intensity, well above the threshold for visibility and 1 kR or more above the darker auroral regions.

Given the unprecedented brightness of the speckles, we carried out a number of tests to rule out potential artefacts, including: repeating the transmission pattern at a different time to rule out radio-frequency (r.f.) interference with the camera electronics, verifying from radar records that no aircraft were in the area, and measuring the periods of white-light sources such as nearby communications towers and airport beacons. We attempted to reproduce the results whenever an aurora moved into range, but auroral events later in the experiment window never produced E layers of sufficient density to support significant transmissions at

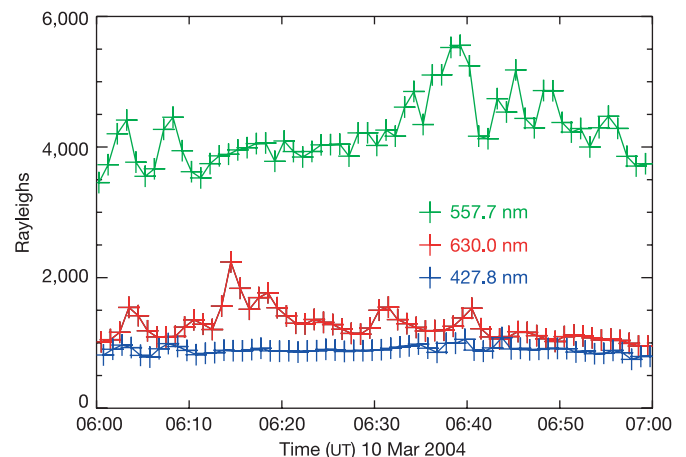


Figure 3 Calibrated measurements from the HAARP imager showing the intensities of various auroral emissions including the oxygen green line, which was well above the threshold for detection by the naked eye during the experiment. The plot covers the period from 6–7 UT on 10 March 2004. The HAARP imager observed an approximately 16° field of view centred on the magnetic zenith and took images at each wavelength once per minute. The red line at 630.0 nm results from the 1D state of neutral atomic oxygen (1.96-eV excitation energy), and is sensitive to lower-energy auroral particles at higher altitudes, while the green line at 557.7 nm arises from the 1S state (4.17 eV) and corresponds to more energetic particles and somewhat lower altitudes³. The 427.8-nm emissions from the first negative band of N_2^+ are an indicator of the ionization rate in the E layer¹¹. At the ~100-km altitude of the E layer, more frequent collisions between neutral molecules deactivate or 'quench' the long-lived 1D state before it can produce the red line (effective lifetime of the order of 1 min)², leaving only the green line with its shorter lifetime of about 0.7 s (ref. 3) in artificial emissions from the E region⁵.

the original frequency again. More detailed analysis of the data revealed additional weak speckles earlier in the original 6-UT hour, at about 06:20 UT, when the transmitter was operated at a lower frequency (and lower effective power), and some of the brightest speckles were also identified in data from one of the other lower-resolution camera systems operated from a separate building, eliminating any doubt that the speckles represent actual light from the sky.

Although visible levels of artificial optical emissions have not been reported previously, there have been other attempts made to stimulate the auroral E layer with radio waves. A similar experiment that used low-light television cameras and a 2-s on-off cycle but different polarization reported an estimated modulation of less than 10 R, interpreted as radio-induced decreases in the green line emission⁷. Large-scale structural changes in the overhead aurora have been reported in conjunction with E-layer heating⁸, but the extremely small number of cases and the close similarity of the observed effects to naturally occurring processes make it difficult to assess the true influence of the radio waves on the auroral events. In contrast, the recent HAARP observations demonstrate clear on-off control of the speckles over 50 or more complete cycles.

Potential sources of the observed bright speckles fall into two categories: production in the local E-region ionosphere by the transmitter beam, or indirect creation by modification of the auroral particle precipitation, which then produces the optical speckles in the same way as the background aurora. If the speckles are locally generated, the role of the natural aurora would probably be limited to creation of the E layer for the radio waves to interact with, and it might be possible to generate similar phenomena in non-auroral E layers independent of any specific on-off cycling, a potentially desirable condition for creation of visible artificial light. If, on the other hand, the speckles result from modification of the auroral particle population, perhaps through perturbations to currents flowing in the E layer or a wave resonance, we expect that the specific frequency of the on-off cycling relative to the natural pulsation frequencies might be a critical parameter, and experiments of this type could potentially become a new tool for exploration of time-dependent processes in the aurora and magnetosphere. □

Received 13 September; accepted 6 December 2004; doi:10.1038/nature03243.

1. Bailey, V. A. On some effects caused in the ionosphere by electric waves, Part II. *Phil. Mag.* **26**(7), 425–453 (1938).
2. Bernhardt, P., Duncan, L. M. & Tepley, C. A. Artificial airglow excited by high-power radio waves. *Science* **242**, 1022–1027 (1988).
3. Chamberlain, J. W. *Physics of the Aurora and Airglow* (International Geophysics Series Vol. 2, Academic, New York, 1961).
4. Rishbeth, H. & Garriot, O. K. *Introduction to Ionospheric Physics* (Academic, New York, 1969).
5. Djuth, F. T. et al. Large airglow enhancements produced via wave-plasma interactions in sporadic E. *Geophys. Res. Lett.* **26**, 1557–1560 (1999).
6. Omholt, A. *The Optical Aurora 6* (Physics and Chemistry in Space Vol. 4, Springer, New York, 1971).
7. Sergienko, T., Kornilov, I., Belova, E., Turunen, T. & Manninen, J. Optical effects in the aurora caused by ionospheric HF heating. *J. Atmos. Sol-Terr. Phys.* **59**, 2401–2407 (1997).
8. Blagoveshchenskaya, N. F. et al. Ionospheric HF pump wave triggering of local auroral activation. *J. Geophys. Res.* **106**, 29071–29089 (2001).
9. Bauer, S. J. *Physics of Planetary Ionospheres* 82–95 (Physics and Chemistry in Space Vol. 6, Springer, New York, 1973).
10. Davis, N. *The Aurora Watchers Handbook* 58 (Univ. Alaska Press, Fairbanks, 1992).
11. Carlson, H. C. Jr & Egeland, A. in *Introduction to Space Physics* (eds Kivelson, M. G. & Russell, V.) 459–502 (Cambridge Press, New York, 1995).

Acknowledgements HAARP is a Department of Defense programme operated jointly by the US Air Force and US Navy. We thank E. Mishin for his contributions to the experiment planning and P. Ning for operating the all-sky imager.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.R.P. (todd.pedersen@hanscom.af.mil).

Observation of random-phase lattice solitons

Oren Cohen¹, Guy Bartal¹, Hrvoje Buljan^{1,2}, Tal Carmon¹, Jason W. Fleischer^{1,3}, Mordechai Segev^{1,3} & Demetrios N. Christodoulides⁴

¹Physics Department and Solid State Institute, Technion – Israel Institute of Technology, Haifa 32000, Israel

²Department of Physics, University of Zagreb, PP331, 10000 Zagreb, Croatia

³Electrical Engineering Department, Princeton University, Princeton, New Jersey 08544, USA

⁴School of Optics/CREOL, University of Central Florida, Florida 32816-2700, USA

The coherence of waves in periodic systems (lattices) is crucial to their dynamics, as interference effects, such as Bragg reflections, largely determine their propagation. Whereas linear systems allow superposition, nonlinearity introduces a non-trivial interplay between localization effects, coupling between lattice sites, and incoherence. Until recently, all research on solitary waves (solitons) in nonlinear lattices has involved only coherent waves. In such cases, linear dispersion or diffraction of wave packets can be balanced by nonlinear effects, resulting in coherent lattice (or ‘discrete’) solitons^{1,2}; these have been studied in many branches of science^{3–11}. However, in most natural systems, waves with only partial coherence are more common, because fluctuations (thermal, quantum or some other) can reduce the correlation length to a distance comparable to the lattice spacing. Such systems should support random-phase lattice solitons displaying distinct features¹². Here we report the experimental observation of random-phase lattice solitons, demonstrating their self-trapping and local periodicity in real space, in addition to their multi-peaked power spectrum in momentum space. We discuss the relevance of such solitons to other nonlinear periodic systems in which fluctuating waves propagate, such as atomic systems, plasmas and molecular chains.

Wave propagation in periodic potentials exhibits universal characteristic features, such as Bragg reflections, bandgaps (stop bands) and alternating regions of normal and anomalous dispersion. In solid-state physics, this leads to regions where the effective mass is positive or negative¹³, whereas in optical settings such as waveguide arrays¹ the same processes correspond to regimes of normal/anomalous diffraction. These effects all result from interference among waves propagating in the lattice, and as such they critically depend on the degree of coherence of the waves involved. The propagation dynamics of incoherent waves in nonlinear periodic potentials depend on the threefold interplay between interference effects, nonlinearity and the statistical (coherence) properties of the waves. As a result, many fundamental physical phenomena occurring with coherent waves, such as solitons^{1–10}, modulation instability^{6,14} and Bloch oscillations^{15,16}, are expected to exhibit new features with incoherent (random-phase) waves. Likewise, potential applications will offer new possibilities when partial coherence plays a role and is used as an additional degree of freedom. For example, using random-phase waves as a statistical probe may provide a valuable tool for exploring general dynamics in nonlinear lattices¹⁷. Another example is the extension of our current experimental findings to obtain Brillouin zone spectroscopy. To date, the only work in this direction has been the theoretical prediction of one-dimensional random-phase lattice solitons (RPLS)¹². We report the experimental observation of incoherent optical lattice solitons in a nonlinear waveguide array.

Before discussing the new features introduced by the partial coherence of waves, we will highlight the most relevant physical

processes associated with the lattice. As periodicity and Bragg reflections dominate the linear properties, and as the spatial power spectrum characterizes the probe beam, it is best to consider the lattice properties through its k -space (momentum space) representation. Figure 1a shows the first and second Brillouin zones of a two-dimensional square lattice with the high-symmetry points (Γ , X and M) marked with white dots. The first two bands in the transmission spectrum are shown in Fig. 1b, and the dispersion curves between the symmetry points of the first two bands are plotted in Fig. 1c. Negative curvature in these figures corresponds to normal diffraction; that is, a narrow beam with its power spectrum centred in a normal diffraction region acquires a convex phase front during linear propagation, in a fashion similar to diffraction in homogeneous media. On the other hand, positive curvature in the dispersion relation corresponds to regions of anomalous diffraction, in which a beam (wave packet) acquires a concave phase front while propagating linearly. A self-focusing nonlinearity can only counteract the broadening tendency of a narrow beam experiencing normal diffraction. Hence, a bright soliton in a self-focusing nonlinear photonic lattice should significantly populate those regions in k -space where the diffraction is normal. Consequently, the power spectrum of random-phase lattice solitons in self-focusing media is expected to have multiple humps, with the humps located in regions of negative (normal) curvature¹². Specifically, for a self-focusing, square, nonlinear photonic lattice, Fig. 1b and c indicates that the power spectrum of a two-band RPLS should consist of a central hump centred on the Γ -point of the first band together with four humps centred on the vicinity of the X-points of the second band.

Our experiments were carried out in optically induced nonlinear photonic lattices¹⁸, recently used for the experimental observations of two-dimensional lattice solitons¹⁹, spatial gap solitons^{19,20}, vortex lattice solitons^{21,22}, dipole-mode discrete solitons²³ and 'optical polarons'²⁴. In this system, plane waves interfering in a photorefractive crystal induce a photonic lattice (Fig. 1d), while voltage applied across the crystal controls the strength of the photorefractive screening nonlinearity^{25,26}. We photograph the light leaving the

output face of the crystal both in real space and in k -space (for the latter we use a lens, and monitor the intensity at the focal plane). A quasi-thermal (partially spatially coherent) quasi-monochromatic source is established by focusing a laser beam onto a rotating diffuser^{27,28}, and then imaging 1:1 with a telescopic (so-called '4f') imaging system onto the input face of the crystal (Fig. 1e). We control the degree of spatial coherence and the power spectrum of the exciting beam by a spatial filter in the Fourier plane of the 4f system. Finally, we note that the nonlinear response time of the medium is much longer than the fluctuation time of the incoherent beam; hence, it can support incoherent solitons^{27,29}.

Figure 2a shows a real-space photograph of the induced square lattice and of a probe beam launched into the lattice, at the input face of the crystal. The lattice period is $\sim 11.5 \mu\text{m}$, and the probe beam has a width of $26 \mu\text{m}$ (full-width at half-maximum, FWHM). Hence, the probe beam covers an area of several (~ 10 – 12) channels. Figure 2b shows the Fourier power spectrum when the probe beam is incoherent (wide blue circle) and of the lattice-forming waves (four dots; 'delta functions'). Because the inducing array beams form Bragg angles in our system, the corresponding square defined by the four dots outlines the first Brillouin zone (Fig. 1a). The wide, blue circle shows that the power spectrum of the incoherent probe beam covers all the first Brillouin zone and significant parts of the second zone. For comparison, Fig. 2c shows the power spectrum of a coherent probe beam (with the diffuser removed) possessing the same size (envelope) in real space as the incoherent probe. The spectral width of such a coherent probe beam is much less than that of the incoherent probe (by a factor greater than 3), and covers only a small portion of the first Brillouin zone. Going back to Fig. 2b, the width indicates that the average speckle size in the incoherent beam is less than one-third the size of the envelope of the beam, and that the spatial correlation distance of this random-phase beam is approximately equal to the lattice period.

The incoherent beam is focused into the lattice with the real-space distribution and power spectrum shown in Fig. 2a and b, respectively, and propagates through a 5-mm nonlinear photonic

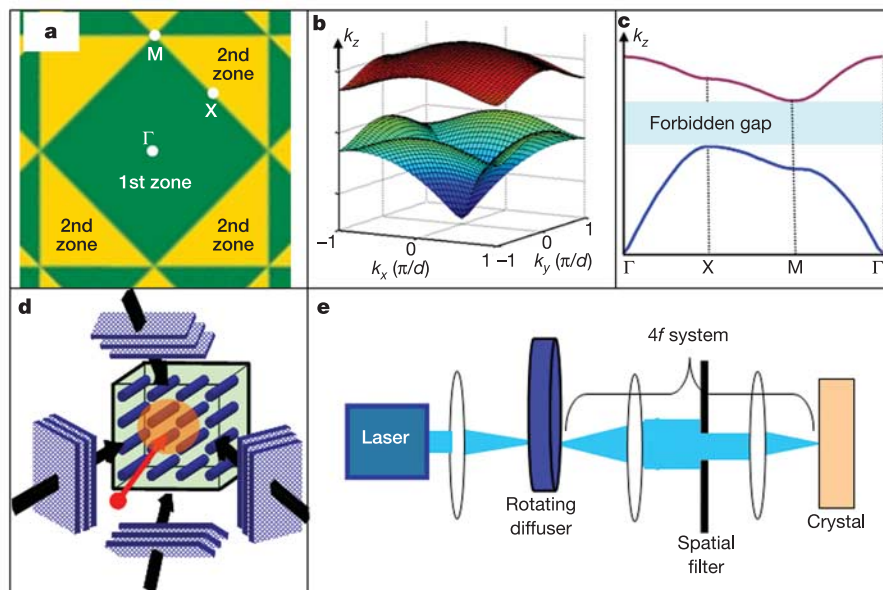


Figure 1 Experimental scheme for producing random-phase solitons in an optically induced square lattice. **a**, First (green square) and second (four yellow triangles) Brillouin zones of a two-dimensional square lattice with the high-symmetry points (Γ , X and M) marked with white dots. **b**, Transmission spectrum of the first two bands of a two-dimensional square lattice with a lattice period d . **c**, Dispersion curves between the symmetry points of the first two bands. Negative curvature in these curves corresponds to

normal diffraction regions. **d**, Diagram of the optical induction technique used to obtain the two-dimensional, square photonic lattice. The blue planes with heavy black arrows indicate the plane waves used to optically induce the lattice. The red arrow indicates the direction of a probe beam entering the lattice. The orange circle indicates the width of the probe beam. **e**, Diagram of our set-up for obtaining a spatially incoherent (quasi-thermal), quasi-monochromatic beam.

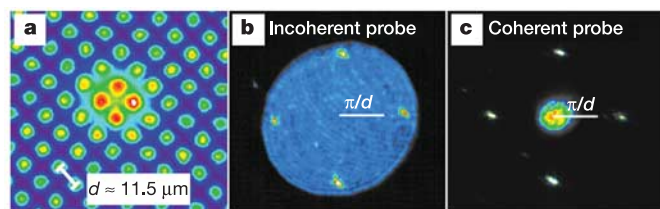


Figure 2 Experimental images at the input face. **a**, Real-space picture of the induced square lattice (with lattice period d) and a probe beam. **b**, Power spectrum of the incoherent probe beam (blue circle) and lattice beams (four dots), which define the corners of the first Brillouin zone. The power spectrum resides in the first and second Brillouin zones. **c**, Same as **b** but with a coherent probe beam that has the same size in real space (as in **a**).

lattice. As in ref. 19, each array beam has ~ 15 mW of power, while the nonlinearity is set by applying 1 kV across the crystal. Typical experimental pictures taken at the output face are shown in Fig. 3. Figure 3a shows linear diffraction for a low-intensity probe beam, characterized by a 1:50 intensity ratio between the probe and array beams. Note that the diffracted output is roughly 2.5 times the input width of the probe. At $10\times$ higher intensity (1:5 ratio), the incoherent probe beam self-traps and forms an RPLS. Figure 3b shows an RPLS centred on a lattice site (a waveguide), while Fig. 3c and d shows RPLSs centred between 2 lattice sites and 4 lattice sites, respectively. Figure 3e depicts the power spectrum of the RPLSs. As before, the four dots represent the array beams, defining the M-symmetry points at the corners of the first Brillouin zone. The power spectrum of the solitons takes on the square symmetry of the lattice and is clearly multi-humped, with well-separated peaks located in the normal diffraction regions of the first two Brillouin zones. For comparison, we monitor the linear and nonlinear propagation of the probe beam with the index lattice removed (that is, in a homogeneous medium). To do this, we block the lattice beams, while maintaining all the other experimental parameters (applied bias field and intensity). We observe that the beam undergoes self-focusing, yet the nonlinearity is too weak to fully balance the linear diffraction broadening of the beam, and thus cannot form an incoherent soliton. Figure 3f shows the power spectrum of the output probe beam. In a sharp contrast to the RPLS spectrum, the self-focused, incoherent beam propagating in the homogeneous (no lattice) nonlinear crystal retains its homogeneity, exhibiting no multi-hump structure.

Next we study the formation of an RPLS with a beam that is more coherent than that of Fig. 3, under similar self-focusing conditions. Figure 4a and b shows respectively the real-space photograph and the power spectrum of the beam, at the input face of the crystal. In this case, the input power spectrum covers (approximately) the first

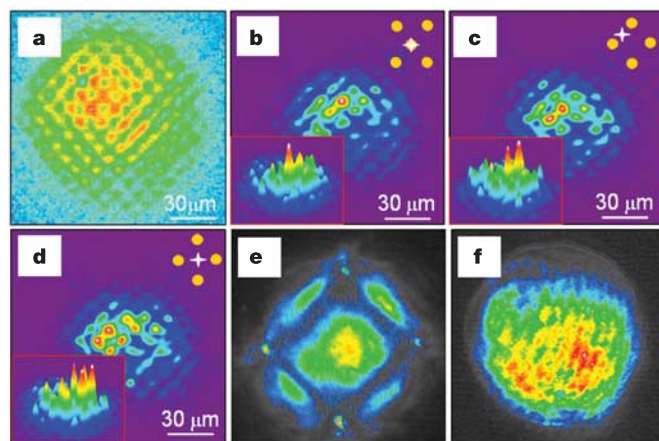


Figure 3 Experimental images at the output face. **a**, Linear lattice diffraction of the incoherent beam. **b–d**, Real-space pictures of random-phase lattice solitons (RPLSs) centred on a site (**b**), between two sites (**c**), and between four sites (**d**). The insets show three-dimensional perspectives of the solitons. Note that **a–d** have the same scale, and the linearly diffracted beam of **a** is at least 2.5 times the width of the soliton. **e**, Power spectrum of an RPLS, displaying its square symmetry and multi-humped structure with the peaks located in the normal diffraction regions of the first two bands. The four dots correspond to the power spectrum of the array beams. **f**, Power spectrum of the self-focused incoherent probe beam at the output face of the crystal without the lattice. Red (blue) colour indicates high (low) intensity.

Brillouin zone only. Figure 4c shows linear diffraction in the lattice for a low-intensity probe beam, while Fig. 4d shows the RPLS when the beam intensity is sufficiently increased. As a consequence of the increased coherence, the linear broadening of the beam is smaller and the formed RPLS is tighter (relative to Fig. 3b). Finally, Fig. 4e depicts the power spectrum of the RPLS. Interestingly, this spectrum is also multi-humped, populating the normal diffraction parts of the first and the second (initially unexcited) Brillouin zones. That is, the input beam with a power spectrum in the first band only reshapes into the RPLS, and in doing so transfers part of the power to the normal diffraction region of the second band while leaving the anomalous diffraction region of the first band unpopulated.

We note that originally, while planning these experiments, we believed that we would have to ‘engineer’ the power spectrum of the initial probe beam to match the multi-humped k -distribution of the RPLS¹². Unexpectedly, we found experimentally that a probe beam with a homogeneous k -space distribution (a single hump with a proper width) self-adjusts its spectrum and evolves naturally, under proper nonlinear conditions, into an RPLS. This cannot be seen using real-space imaging, but the Fourier imaging technique

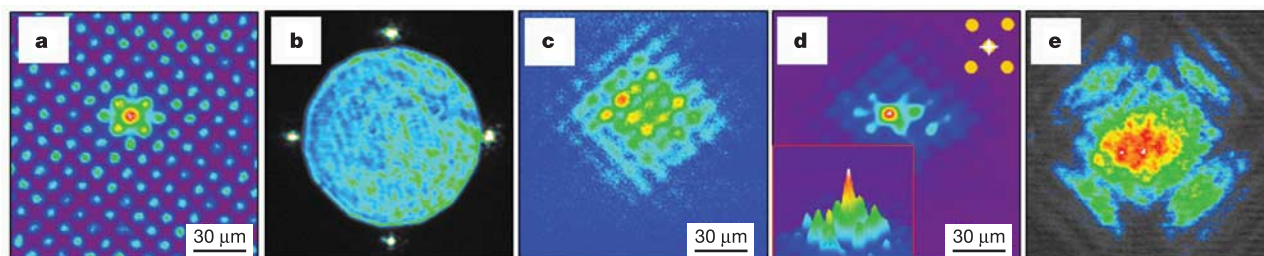


Figure 4 An RPLS with a lower incoherence than the RPLS of Figs 2 and 3. **a**, Real-space photograph of induced square lattice and the probe beam. **b**, Power spectrum of the incoherent input probe beam (blue circle) and lattice beams (four dots); the input spectrum

resides in the first band only. **c**, Linear lattice diffraction of the incoherent beam. **d, e**, Real-space image (**d**) and power spectrum (**e**) of the emerging RPLS under proper self-focusing conditions.

presented here clearly shows the difference between the simple spectrum of the input (Fig. 2b) and the complex multi-hump spectrum of the output (Fig. 3e). During propagation, then, there is a k -space evolution in which energy transfers from regions of anomalous diffraction to regions of normal diffraction. Moreover, Fig. 4b and e clearly shows that energy transfers between bands. Because different k -vectors correspond to different Bloch modes, this energy transfer corresponds to an inherently nonlinear coupling between modes of the linear lattice. In this sense, our measurements are related to the Fermi–Pasta–Ulam problem³⁰. In the Fermi–Pasta–Ulam system, a lattice mode was excited, and it was expected that nonlinearity would redistribute the energy to a homogeneous state, that is, lead to equipartition. Instead, energy recurred to the initial mode, exhibiting energy oscillations between the initial mode and a finite group of modes, in an almost periodic fashion. In our system, an initially homogeneous k -space distribution evolves into a steady-state multi-humped soliton power spectrum. The focus of this Letter has been on the observation of these RPLs, but the rich dynamics underlying this energy transfer suggests many applications. For example, the spatially incoherent input beam can be used as a probe with given statistics, and the imaging techniques outlined here allow the observation of nonlinear effects, in both real and Fourier space.

Our experiments open up new possibilities in other nonlinear periodic systems beyond optics. For example, one can think of random-phase matter-wave lattice solitons in Bose–Einstein condensates, where the periodic potential is also optically induced. Similarly, one can envision random-phase solitons occurring with vibrational waves propagating along periodic molecular structures, finite-temperature plasma waves, or with charge-density waves in polymers or in crystalline conductors. In a more general sense, RPLs should exist in any nonlinear periodic system, because fluctuations (quantum, thermal, and so on) are always present and the propagating waves are never fully correlated. □

Received 24 June; accepted 10 December 2004; doi:10.1038/nature03267.

- Christodoulides, D. N., Lederer, F. & Silberberg, Y. Discretizing light behaviour in linear and nonlinear waveguide lattices. *Nature* **424**, 817–823 (2003).
- Campbell, D. K., Flach, S. & Kivshar, Yu. S. Localizing energy through nonlinearity and discreteness. *Phys. Today* **57**(1), 43–49 (Jan. 2004).
- Davydov, A. S. & Kislukha, N. I. Solitary excitation in one-dimensional molecular chains. *Phys. Status Solidi B* **59**, 465–470 (1973).
- Xie, A., van der Meer, L., Hoff, W. & Austin, R. H. Long-lived amide I vibrational modes in myoglobin. *Phys. Rev. Lett.* **84**, 5435–5438 (2000).
- Su, W. P., Schieffer, J. R. & Heeger, A. J. Solitons in polyacetylene. *Phys. Rev. Lett.* **42**, 1698–1701 (1979).
- Christodoulides, D. N. & Joseph, R. I. Discrete self focusing in nonlinear arrays of coupled waveguides. *Opt. Lett.* **13**, 794–796 (1988).
- Eisenberg, H. S., Silberberg, Y., Morandotti, R., Boyd, A. R. & Aitchinson, J. S. Discrete spatial optical solitons in waveguide arrays. *Phys. Rev. Lett.* **81**, 3383–3386 (1998).
- Sievers, A. J. & Takeno, S. Intrinsic localized modes in anharmonic crystal. *Phys. Rev. Lett.* **61**, 970–973 (1988).
- Schwartz, U. T., English, L. Q. & Sievers, A. J. Experimental generation and observation of intrinsic localized spin wave modes in an antiferromagnet. *Phys. Rev. Lett.* **83**, 223–226 (1999).
- Trombettoni, A. & Smerzi, A. Discrete solitons and breathers with dilute Bose–Einstein condensate. *Phys. Rev. Lett.* **86**, 2353–2356 (2001).
- Eiermann, B. et al. Bright Bose–Einstein gap solitons of atoms with repulsive interaction. *Phys. Rev. Lett.* **92**, 230401 (2004).
- Buljan, H. et al. Random phase solitons in nonlinear periodic lattices. *Phys. Rev. Lett.* **92**, 223901 (2004).
- Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* (Saunders, Philadelphia, 1976).
- Meier, J. et al. Experimental observation of discrete modulation instability. *Phys. Rev. Lett.* **92**, 163902 (2004).
- Pertsch, T., Dannberg, P., Elflein, W., Brauer, A. & Lederer, F. Optical Bloch oscillations in temperature tuned waveguide arrays. *Phys. Rev. Lett.* **83**, 4752–4755 (1999).
- Morandotti, R., Peschel, U., Aitchison, J. S., Eisenberg, H. S. & Silberberg, Y. Experimental observation of linear and nonlinear optical Bloch oscillations. *Phys. Rev. Lett.* **83**, 4756–4759 (1999).
- Frisch, U. *Turbulence* (Cambridge Univ. Press, Cambridge, UK, 1995).
- Efremidis, N. K., Sears, S., Christodoulides, D. N., Fleischer, J. W. & Segev, M. Discrete solitons in photorefractive optically-induced photonic lattices. *Phys. Rev. E* **66**, 046602 (2002).
- Fleischer, J. W., Segev, M., Efremidis, N. K. & Christodoulides, D. N. Observation of two-dimensional discrete solitons in optically induced nonlinear photonic lattices. *Nature* **422**, 147–150 (2003).
- Fleischer, J. W., Carmon, T., Segev, M., Efremidis, N. K. & Christodoulides, D. N. Observation of discrete solitons in optically-induced real time waveguide arrays. *Phys. Rev. Lett.* **90**, 023902 (2003).

- Fleischer, J. W. et al. Observation of vortex ring “discrete” solitons in 2D photonic lattices. *Phys. Rev. Lett.* **92**, 123904 (2004).
- Neshev, D. N. et al. Observation of discrete vortex solitons in optically-induced photonic lattices. *Phys. Rev. Lett.* **92**, 123903 (2004).
- Neshev, D., Ostrovskaya, E., Kivshar, Yu. S. & Krokowski, W. Spatial solitons in optically induced gratings. *Opt. Lett.* **28**, 710–712 (2003).
- Martin, H., Eugenieva, E. D., Chen, Z. G. & Christodoulides, D. N. Discrete solitons and soliton-induced dislocations in partially coherent photonic lattices. *Phys. Rev. Lett.* **92**, 123902 (2004).
- Segev, M., Valley, G. C., Crosignani, B., Diporto, P. & Yariv, A. Steady state spatial screening-soliton in photorefractive media with external applied field. *Phys. Rev. Lett.* **73**, 3211–3214 (1994).
- Christodoulides, D. N. & Carvalho, M. I. Bright, dark and gray spatial soliton states in photorefractive media. *J. Opt. Soc. Am. B* **12**, 1628–1633 (1995).
- Mitchell, M., Chen, Z., Shih, M. & Segev, M. Self-trapping of partially spatially-incoherent light. *Phys. Rev. Lett.* **77**, 490–493 (1996).
- Goodman, J. *Statistical Optics* (Wiley & Sons, New York, 1985).
- Mitchell, M. & Segev, M. Self-trapping of incoherent white light. *Nature* **387**, 880–883 (1997).
- Fermi, E., Pasta, J. & Ulam, S. Report LA-1940 (LANL, Los Alamos, 1955).

Acknowledgements This work was supported by the Israeli Science Foundation, the Israel–USA Binational Science Foundation, and by the German–Israeli DIP Project. O.C. acknowledges the generous support of the Israeli Ministry of Science through the Eshkol Fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S. (msegev@tx.technion.ac.il).

Conversion of large-amplitude vibration to electron excitation at a metal surface

Jason D. White¹, Jun Chen¹, Daniel Matsiev¹, Daniel J. Auerbach² & Alec M. Wodtke¹

¹Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, USA

²Hitachi Global Storage Technologies, 650 Harry Road, San Jose, California 95120-6099, USA

Gaining insight into the nature and dynamics of the transition state is the essence of mechanistic investigations of chemical reactions¹, yet the fleeting configuration when existing chemical bonds dissociate while new ones form is extremely difficult to examine directly². Adiabatic potential-energy surfaces—usually derived using quantum chemical methods³ that assume mutually independent nuclear and electronic motion⁴—quantify the fundamental forces between atoms involved in reaction and thus provide accurate descriptions of a reacting system as it moves through its transition state^{5,6}. This approach, widely tested for gas-phase reactions⁷, is now also commonly applied to chemical reactions at metal surfaces⁸. There is, however, some evidence calling into question the correctness of this theoretical approach for surface reactions: electronic excitation upon highly exothermic chemisorption has been observed⁹, and indirect evidence suggests that large-amplitude vibrations of reactant molecules can excite electrons at metal surfaces^{10,11}. Here we report the detection of ‘hot’ electrons leaving a metal surface as vibrationally highly excited NO molecules collide with it. Electron emission only occurs once the vibrational energy exceeds the surface work function, and is at least 10,000 times more efficient than the emissions seen in similar systems where large-amplitude vibrations were not involved^{12–18}. These observations unambiguously demonstrate the direct conversion of vibrational to electronic excitation, thus questioning one of the basic assumptions currently used in theoretical approaches to describing bond-dissociation at metal surfaces.

Electronically adiabatic potential-energy surfaces computed with *ab initio* quantum chemistry methods have played a central role in advancing our understanding of chemical reactions. This powerful conceptual and predictive system is based on the assumption that nuclear motion associated with the reorganization of chemical bonds proceeds without facile electronic excitation (called the electronically adiabatic or Born–Oppenheimer approximation). Although this assumption has been widely tested for gas-phase reactions, less is known for reactions at metal interfaces.

Perhaps the greatest value of the electronically adiabatic potential-energy surface stems from its use in providing an accurate description of motion through a chemical transition state, given that experimental detection of chemical transition states is extremely difficult (see ref. 2 for a notable exception). Thus for most reactions, theory plays an essential role in revealing their fundamental nature.

Motion through a transition state involves large-amplitude vibration: bond elongation is so extreme that dramatic changes occur in the structure of the electronic cloud holding the atoms together as the system reorganizes to new chemical species. There is already indirect evidence that under these conditions, heavy-atom vibrational motion can exchange large amounts of energy, producing electronic excitations in a metal, which casts doubt on the electronically adiabatic picture of chemical transition states for reactions at metal interfaces^{10,11}. As yet, however, there are no direct observations of ‘hot’ electrons produced by excitation exchange with large-amplitude vibration.

Here we report experiments intended to demonstrate the importance of non-adiabatic electronic effects associated with large-amplitude intramolecular motions similar to those associated with transition-state traversal. To carry out this work, we used an apparatus similar to that of ref. 19. A pulsed molecular beam containing NO with a kinetic energy of 29 meV is formed in the ‘source’ vacuum chamber. It then passes through a skimmer into another ‘excitation and detection’ vacuum chamber, where it is resonantly excited with laser light to prepare selected quantum states in high vibrational levels. This chamber also provides a laser-induced fluorescence (LIF) detection station for monitoring the intensity of the beam of vibrationally excited molecules. After passing through a 1-mm aperture into an ultrahigh-vacuum surface science chamber, the molecules collide with the metal surface,

inducing emission of electrons. Stimulated emission pumping²⁰, SEP(ν), was used to produce NO molecules in a selected excited vibrational states, ν , where $18 \geq \nu \geq 4$. This was accomplished through excitation of NO to the excited electronic state $A^2\Sigma^+(\nu' = 3)$ followed by laser-induced de-excitation to the targeted vibrational state, ν , back in the ground ($X^2\Pi$) electronic state. We also used spontaneous emission from laser-prepared $A^2\Sigma^+(\nu = 0)$ to populate vibrationally excited states, $X^2\Pi(\nu \leq 5)$.

To obtain the relative electron emission signals for different vibrational states, we compared the signal produced by the SEP-prepared state to the residual signal produced in the absence of the de-excitation laser, where vibrational states are populated by spontaneous emission from $A^2\Sigma^+(\nu' = 3)$. This provided an internal standard of electron emission that allowed us to gain a measure of control over the day-to-day reproducibility of the experiment.

To determine the per-collision efficiency for electron emission, the absolute number of NO(ν) molecules in the molecular beam pulses was determined from saturated laser-induced fluorescence detection at the position where the beam collides with the surface. The absolute number of fluorescence photons was derived from the observed current of a solar-blind photomultiplier tube, accounting for: (1) measured photomultiplier gain, (2) photocathode quantum efficiency, (3) optical collection efficiency of the lens system, and (4) spatial overlap of the laser and molecular beams. Emitted electrons were counted with a multi-channel scaler.

We prepared low-work-function surfaces by depositing Cs on Au(111) with a commercially available source (SAES Getters) in a ultrahigh-vacuum chamber with base pressure of $\sim 1 \times 10^{-10}$ torr. The surface temperature was ~ 300 K in all the experiments reported here. The work function of various metals with submonolayer Cs

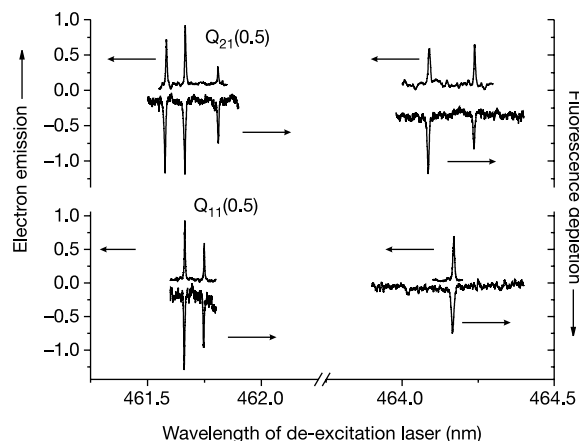


Figure 1 Electron emission from NO($\nu = 18$) collisions with Cs/Au(111) surface. Electron emission (up-going signal) is detected as a function of the de-excitation laser's wavelength for two excitation transitions ($Q_{21}(0,5)$ and $Q_{11}(0,5)$). We compare these spectra to fluorescence depletion (down-going signal) spectra observed under identical conditions. The observed spectral resonances agree with known transitions of the $A^2\Sigma^+(\nu' = 3)$ to $X^2\Pi(\nu = 18)$ system of NO to better than the line-width of the laser.

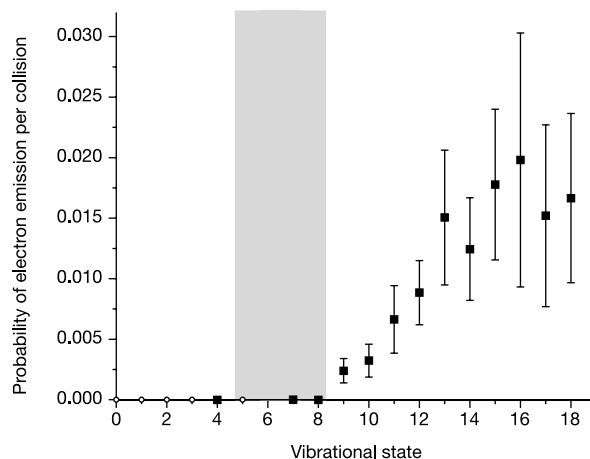


Figure 2 Vibrational dependence of electron emission efficiency. Stimulated emission pumping (solid squares) and spontaneous emission from the $A^2\Sigma^+(\nu = 0)$ state (open circles) lead to relative electron emission probabilities for the indicated vibrational states. The points are average values of at least five measurements for each value of ν . The error bars reflect an analysis of error propagation from measured uncertainties in factors contributing to the derived quantum yield. These include uncertainties in excitation and de-excitation efficiency. These error bars are similar to 90% confidence intervals (using Student's *t*-test) derived from statistical analysis of multiple measurements. For $\nu = 13$ and 16, laser de-excitation is accompanied by some absorption out of the excited electronic state. This does not contribute to the electron emission signal but increases the error in the measurement. Experiments producing $\nu = 6$ were not technically feasible. Error bars for data indicated by open circles are smaller than the symbol. The shaded region indicates the vibrational equivalent of the surface work function ($\nu = 6$ and 7 correspond to 1.34 and 1.55 eV, respectively).

coverage drops below the asymptotic work functions of both the base metal and pure Cs^{17,21–23}. We monitored the effect of Cs dosage on the work function by observing the photoemission generated by a He-Ne laser ($h\nu = 1.96$ eV), which enables us to prepare surfaces that exhibit maximum photoemission and thus a surface near the work function minimum²². Cs-covered gold surfaces prepared in this way have work functions between 1.3 and 1.6 eV (ref. 21).

We detected emitted electrons with a 40-mm double micro-channel-plate detector located close to the surface and identified the observed signals as electrons by the fact that a 2–3-gauss magnetic field effectively removes the signal. The angle of incidence between the molecular beam and the surface normal was fixed at 60° to maximize electron collection.

Figure 1 shows the results observed when monitoring the electron current emitted from the surface, while scanning the de-excitation laser wavelength (upward-going signals) in the SEP ($\nu = 18$) pumping scheme. Resonances in the depletion of side fluorescence²⁰ were measured simultaneously at the preparation zone. These are also shown in Fig. 1 (downward-going signals). There is a one-to-one correspondence between electron emission and fluorescence depletion resonances, establishing that incident NO ($\nu = 18$) is responsible for the observed electron signal. The observed resonance wavelengths can be predicted from spectroscopic theory of the NO molecule to better than the bandwidth of the laser (0.12 cm⁻¹).

Collisions of NO (ν) with low-work-function surfaces resulted in electron emission for $\nu = 9$ –18. States higher than $\nu = 18$ were not studied. Tuning the laser away from either of the two resonances (excitation and de-excitation) used for state preparation resulted in a reduction or complete loss of electron emission. SEP was also used to prepare NO ($\nu = 4, 5$, and 7), resulting in no detectable electron emission. NO ($\nu = 8$) was prepared and sometimes yielded a non-zero emission signal. Spontaneous emission from NO- $A^2\Sigma^+$ ($\nu = 0$), which produces molecules in vibrational states $\nu \leq 5$, also gave no electron emission. No electron emission was observed for NO in its ground vibrational state. No surface oxide was required to observe the electron emission, indicating a mechanism of electron emission different from previous reports of exoelectron emission resulting from NO ($\nu = 0$) interactions at surfaces^{16–18}.

The vibrational dependence of the per-collision electron emission efficiency is shown in Fig. 2. For comparison, the energetic equivalent position of the surface work function is shown as a grey-shaded bar. We see a vibrational threshold near the energetic equivalent of the surface work function, strong evidence that we are witnessing direct conversion of vibrational energy to electronic excitation in the metal. These results indicate a yield of approximately 0.02 electrons per incident NO ($\nu = 18$) molecule. By comparison, we note that the yield previously observed for exoelectron emission from ground vibrational state NO, O₂, N₂O and NO₂ on Cs and Li films is of the order of 10⁻⁸ to 10⁻⁶ (refs 12–18).

A detailed mechanistic explanation for the experimental observation of electron emission induced by large-amplitude molecular vibration will require interactions between theory and additional experimental work, but some aspects of the dynamics already seem clear. First, it seems unlikely that we are witnessing electron emission resulting from NO dissociation. Greber¹² has described how a 'hole' on a newly formed O⁻ ion formed after dissociative electron transfer may plummet in energy far below the Fermi level before an additional electron is transferred to produce the ground electronic state O²⁻. The quenching of this hole can result in an Auger emission event. In such an O⁻ 'chemical hole diving' process, electron emission efficiencies of 10⁻⁶–10⁻⁸ were reported^{15,16}, whereas here we observed efficiencies as high as 0.02. Furthermore, such a dissociative mechanism could not explain the observed vibrational threshold, without invoking a coincidental dissociation barrier equal to the work function.

It also seems unlikely that the vibrational energy transfer proceeds in a fashion where single vibrational quanta are transferred efficiently but sequentially in multiple steps. In this case, the vibrational energy would transfer in small amounts to many electrons and it is unlikely that any single electron would acquire energy beyond the work function. Indeed, the observation of a vibrational energy threshold for electron emission near the value of the surface work functions indicates that multiple vibrational quanta are coupled to a single electron, a conclusion that is easily reconciled with a mechanism where an electron is transferred transiently to the NO molecule.

Thus, the considerations presented above strongly suggest that the observed electron emission results from a process that converts large-amplitude vibrational energy of heavy nuclei in the vicinity of the metal surface to translational energy of a metal electron by a mechanism involving electron transfer from the metal to the NO molecule. The observation that large-amplitude vibrational motion can efficiently excite metallic electrons by more than 1.3 eV points to the need to account for strong electronically non-adiabatic influences in theories of chemical reactions at metal interfaces. □

Received 28 September; accepted 18 November 2004; doi:10.1038/nature03213.

1. Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **3**, 107 (1935).
2. Manolopoulos, D. E. *et al.* The transition-state of the F+H₂ reaction. *Science* **262**, 1852–1855 (1993).
3. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, 1133–1138 (1965).
4. Born, M. & Oppenheimer, E. Born Oppenheimer approximation. *Ann. Phys.* **84**, 457–484 (1927).
5. Eyring, H. & Polanyi, M. Über einfache Gasreaktionen. *Sonderdruck aus Z. Phys. Chem.* (special issue) **B12**, 279–311 (1931); Hettner, H. (trans.) in *Quantum Chemistry Classic Scientific Papers* (World Scientific Series in 20th Century Chemistry Vol. 8, Imperial College Press, River Edge, NJ, 2000).
6. Eyring, H., Gershinowitz, H. & Sun, C. E. H + H₂ potential surface. *J. Chem. Phys.* **3**, 786 (1935).
7. Chao, S. D. *et al.* A fully state- and angle-resolved study of the H+HD→D+H₂ reaction: Comparison of a molecular beam experiment to *ab initio* quantum reaction dynamics. *J. Chem. Phys.* **117**, 8341–8361 (2002).
8. Kroes, G. J., Gross, A., Baerends, E. J., Scheffler, M. & McCormack, D. A. Quantum theory of dissociative chemisorption on metal surfaces. *Acc. Chem. Res.* **35**, 193–200 (2002).
9. Nienhaus, H. Electronic excitations by chemical reactions on metal surfaces. *Surf. Sci. Rep.* **45**, 3–78 (2002).
10. Huang, Y. H., Rettner, C. T., Auerbach, D. J. & Wodtke, A. M. Vibrational promotion of electron transfer. *Science* **290**, 111–114 (2000).
11. Diekhoner, L. *et al.* Indirect evidence for strong nonadiabatic coupling in N₂ associative desorption from and dissociative adsorption on Ru(0001). *J. Chem. Phys.* **117**, 5018–5030 (2002).
12. Greber, T. Chemical hole diving. *Chem. Phys. Lett.* **222**, 292–296 (1994).
13. Brandt, M., Greber, T., Bowering, N. & Heinzmann, U. The role of molecular state and orientation in harpooning reactions: N₂O on Cs/Pt(111). *Phys. Rev. Lett.* **81**, 2376–2379 (1998).
14. Brandt, M., Kuhlmann, F., Greber, T., Bowering, N. & Heinzmann, U. Interaction of gas-phase oriented N₂O with lithium metal: evidence for an Eley-Rideal mechanism. *Surf. Sci.* **439**, 49–58 (1999).
15. Bottcher, A. & Giessel, T. Dissociative chemisorption of N₂O molecules on Cs layers monitored via exoelectron emission. *Surf. Sci.* **408**, 212–222 (1998).
16. Bottcher, A., Grobecker, R., Greber, T. & Ertl, G. Negative particle-emission from a Cs/Ru(0001) surface during exposure to NO and NO₂. *Chem. Phys. Lett.* **208**, 404–408 (1993).
17. Bottcher, A., Grobecker, R., Imbeck, R., Morgante, A. & Ertl, G. Exoelectron emission during oxidation of Cs films. *J. Chem. Phys.* **95**, 3756–3766 (1991).
18. Grobecker, R., Greber, T., Bottcher, A. & Ertl, G. Thermally activated emission of exoelectrons accompanying the oxidation of Cs films. *Phys. Status Solidi A Appl. Res.* **146**, 259–267 (1994).
19. Chen, J., Matsiev, D., White, J. D., Murphy, M. & Wodtke, A. M. Hexapole transport and focusing of vibrationally excited NO molecules prepared by optical pumping. *Chem. Phys.* **301**, 161–172 (2004).
20. Silva, M., Jongma, R., Field, R. W. & Wodtke, A. M. The dynamics of "stretched molecules": Experimental studies of highly vibrationally excited molecules with stimulated emission pumping. *Annu. Rev. Phys. Chem.* **52**, 811–852 (2001).
21. Skottke, M. *et al.* Preparation and characterization of thin CsAu films. *Thin Solid Films* **203**, 131–145 (1991).
22. Wang, C. S. High photoemission efficiency of submonolayer cesium-covered surfaces. *J. Appl. Phys.* **48**, 1477–1479 (1977).
23. Woratschek, B., Sesselmann, W., Kuppers, J., Ertl, G. & Haberland, H. The interaction of cesium with oxygen. *J. Chem. Phys.* **86**, 2411–2422 (1987).

Acknowledgements We thank J. Brauman for reviewing an early version of this manuscript. The work was partially supported by a grant from the National Science Foundation as well as a grant from the Department of Energy Office of Basic Energy Sciences.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.M.W. (wodtke@chem.ucsb.edu).

Low European methyl chloroform emissions inferred from long-term atmospheric measurements

Stefan Reimann¹, Alistair J. Manning², Peter G. Simmonds³, Derek M. Cunnold⁴, Ray H. J. Wang⁴, Jinlong Li⁴, Archie McCulloch³, Ronald G. Prinn⁵, Jin Huang⁵, Ray F. Weiss⁶, Paul J. Fraser⁷, Simon O'Doherty³, Brian R. Grealley³, Konrad Stemmler¹, Matthias Hill¹ & Doris Folini¹

¹Empa, Institute of Materials Science and Technology, Ueberlandstrasse 129, 8600 Duebendorf, Switzerland

²Climate Research, Met Office, FitzRoy Road, Exeter EX1 3PB, UK

³School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, UK

⁴School of Earth and Atmospheric Sciences, Georgia Institute of Technology, 221 Bobby Dodd Way, Atlanta, Georgia 30332-0340, USA

⁵Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307, USA

⁶Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0244, USA

⁷Atmospheric Research, Commonwealth Scientific and Industrial Research Organisation, PMB, Aspendale, Victoria 3195, Australia

Methyl chloroform (CH_3CCl_3 , 1,1,1-trichloroethane) was used widely as a solvent before it was recognized to be an ozone-depleting substance and its phase-out was introduced under the Montreal Protocol¹. Subsequently, its atmospheric concentration has declined steadily^{2–4} and recent European methyl chloroform consumption and emissions were estimated to be less than 0.1 gigagrams per year^{1,5}. However, data from a short-term tropospheric measurement campaign (EXPORT) indicated that European methyl chloroform emissions could have been over 20 gigagrams in 2000 (ref. 6), almost doubling previously estimated global emissions^{1,4}. Such enhanced emissions would significantly affect results from the CH_3CCl_3 method of deriving global abundances of hydroxyl radicals (OH) (refs 7–12)—the dominant reactive atmospheric chemical for removing trace gases related to air pollution, ozone depletion and the greenhouse effect. Here we use long-term, high-frequency data from Mace Head, Ireland and Jungfraujoch, Switzerland, to infer European methyl chloroform emissions. We find that European emission estimates declined from about 60 gigagrams per year in the mid-1990s to 0.3–1.4 and 1.9–3.4 gigagrams per year in 2000–03, based on Mace Head and Jungfraujoch data, respectively. Our European methyl chloroform emission estimates are therefore higher than calculated from consumption data^{1,5}, but are considerably lower than those derived from the EXPORT campaign in 2000 (ref. 6).

Consumption of CH_3CCl_3 is to be phased out by 2015. Consequently, global emissions of CH_3CCl_3 decreased from about 720 to 20 Gg (1990–2000; ref. 1) and atmospheric background concentrations have declined from 140–150 parts per trillion (p.p.t.) in the 1990s^{2,3,7,8} to 25 p.p.t. in 2003. However, in modelling measurements of four flights in a campaign in 2000 (EXPORT), European CH_3CCl_3 emissions were estimated to be $>20 \text{ Gg yr}^{-1}$ (ref. 6). This was supported by high concentrations of CH_3CCl_3 measured in a three-week campaign (MINOS) in Crete in 2001 (ref. 13). These large European emissions would more than double the reported global CH_3CCl_3 emissions of about 20 Gg in 2000 (ref. 4) and would significantly impact on estimations of global OH-radical abundances, using global emissions of CH_3CCl_3 (refs 8, 11, 12).

Our study investigates this inconsistency using long-term CH_3CCl_3 measurements at Mace Head (Ireland, 1994–2003) and Jungfraujoch (Switzerland, 2000–2003). Compared to earlier estimates from Mace Head alone^{14,15}, the Jungfraujoch data extend the

spatial coverage towards central and southern Europe. European emission estimates (EU-15 and Switzerland) are derived from inter-species correlations with carbon monoxide (CO) at Mace Head and Jungfraujoch, or from a meteorological transport model (for Mace Head only).

The derived emissions are representative of regions which predominantly influence air masses passing over both sites, providing a good coverage of sources from northwestern and central-western Europe (Ireland, the UK, Belgium, Netherlands, Germany, France) and from northern Italy. Air masses advected from southernmost Europe (Greece, Turkey, Spain, southern Italy) are observed only occasionally and so the data is less representative for these regions.

For the estimation of European emissions by inter-species correlation, values of CO and CH_3CCl_3 above the baseline (excess concentration) were used to detect polluted European air masses arriving at the sites. Baseline concentrations were defined by two independent methods (see Methods): method a is a statistical method and method b is a method based upon the measurement precision (Fig. 1). For both methods, the baseline has been subtracted and annual averages of the excess concentrations of CO and CH_3CCl_3 were calculated. For the estimation of European emissions of CH_3CCl_3 , the annual ratio of the excess concentrations of $\text{CH}_3\text{CCl}_3/\text{CO}$ was multiplied by the annual European emission estimates for CO¹⁶ (see Methods).

A basic assumption of this technique is that the CO and CH_3CCl_3 sources are located together. Although this criterion is not strictly met in reality, it is reasonable to presume that both CH_3CCl_3 and CO emissions are caused by many diffuse, spatially similar anthropogenic sources. This is supported by the fact that at Mace Head (1998–2003) over 80% and at Jungfraujoch (2000–2003) over 50% of all CH_3CCl_3 polluted events are also polluted by CO. However, the accuracy of the estimates decreases for distant regions and underestimation of strong but remote point sources is possible if high emissions of CO are picked up during transport. The use of annual ratios of CO and CH_3CCl_3 during pollution events ensures that estimates are approximately representative of average conditions over Europe.

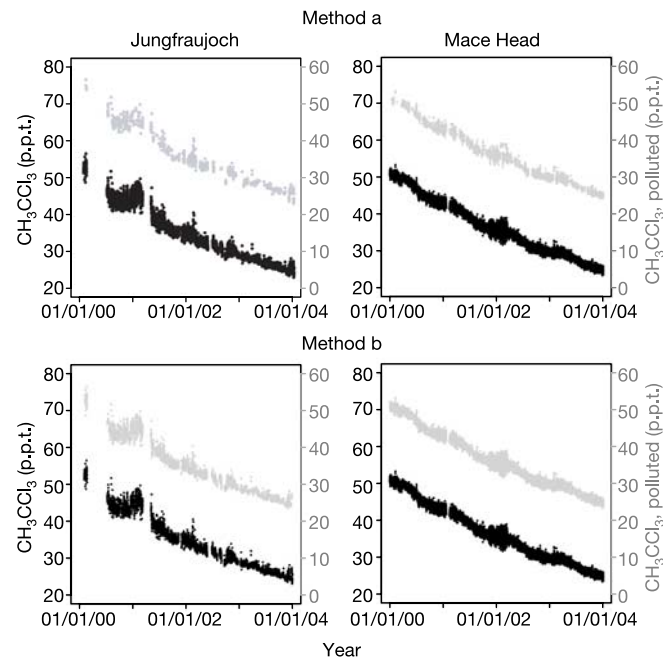


Figure 1 CH_3CCl_3 at Mace Head and Jungfraujoch in 2000–03. The total CH_3CCl_3 data are depicted in black, whereas data collected during CO and CH_3CCl_3 -identified pollution events (detected by methods a and b) are indicated in grey (with a vertical-scale shift of 20 p.p.t. for better visualization).

Table 1 Estimated European emissions of CH_3CCl_3 [Gg yr^{-1}]

Year	Mace Head				Jungfraujoch		Consumption-based
	Method a	Method b	NAME 1-yr	NAME 3-yr	Method a	Method b	
1994	61.1						60.2
1995	20.8		19.1				70.1
1996	10.1		17.8	18.9			22.3
1997	7.4		6.2	11.2			0.06
1998	3.0		2.4	5.5			0.05
1999	1.4		1.6	1.5			0.01
2000	1.0	0.6	1.1	0.9	3.9	2.5	0.02
2001	0.2	0.2	1.7	0.9	4.3	3.4	
2002	0.7	0.1	1.8	1.5	2.7	1.1	
2003	0.5	0.2	1.0		2.8	0.7	
2000–2003	0.6	0.3	1.4	1.1	3.4	1.9	

All values are in gigagrams per year. Methods a and b are the inter-species correlation techniques used. NAME represents the modelled one-year and three-years average values. The consumption-based figures represent the European fraction (using reported consumption) of the global emissions¹.

At Mace Head, European CH_3CCl_3 emissions estimates were about 60 Gg in 1994 and declined to 0.2–1.0 Gg yr^{-1} (method a) or 0.1–0.6 Gg yr^{-1} (method b) in 2000–03 (Table 1). At Jungfraujoch, the annual emission estimates using methods a and b in 2000–03 were 2.7–4.3 and 0.7–3.4 Gg yr^{-1} , respectively.

For Mace Head, the lagrangian particle dispersion model NAME^{14,15} and a best-fit algorithm were used to estimate European emissions (see Methods). One-yearly and three-yearly emission estimates (Table 1) are given by the sums of regionally distributed emissions (Supplementary Table 1). The same sharp decline in emissions is observed as with the inter-species correlation method. The NAME estimates flatten off after 2000 at 1.4 and 1.1 Gg yr^{-1} with the one-yearly and three-yearly estimates, respectively.

The best estimates of 2000–03 European CH_3CCl_3 emissions were calculated separately for both sites, yielding 0.8 Gg yr^{-1} for Mace Head (average of all estimates) and 2.7 Gg yr^{-1} for Jungfraujoch (averaged methods a and b).

European CH_3CCl_3 emissions have also been estimated using consumption data reported under the Montreal Protocol⁵. A delay between consumption and emission of less than one year is applied, as CH_3CCl_3 was mainly used to replace emitted material and stockpiling was not significant, at least before sales were banned in Europe in 1996 (ref. 1). Estimated emissions show a pronounced decrease from 160 Gg in 1990 to <0.1 Gg in 2000 (refs 1, 5) (Table 1 and Supplementary Table 2).

Compared to estimates from Mace Head, consumption-based figures were higher in 1995 but were consistently lower afterwards. This discrepancy could be caused, in part, by consumers banking CH_3CCl_3 , anticipating the restrictions of the Montreal Protocol. This hypothesis is supported by the drastic decline in consumption-based estimates between 1995 and 1996, which is not represented by observation-based estimates. Additional sources of CH_3CCl_3 not covered by industry or Montreal Protocol consumption data could include hazardous waste sites⁶, illegal import and small releases during HCFC production.

When the regional European emissions estimated from the EXPORT data are run in the NAME model for Mace Head, the modelled pollution episodes in 2000 are significantly larger than observed (Fig. 2), suggesting a significant over-estimation of the emissions by EXPORT⁶. For example, in December 2000 a pollution episode of about +18 p.p.t. was modelled, which was considerably above the observations (+1.5 p.p.t.). This overestimate resulted from releasing 9 Gg yr^{-1} of CH_3CCl_3 in Italy (that is, regional emissions estimated from EXPORT⁶). This indicates that Italian sources are probably not increased compared with the rest of Europe, and that southern European sources can occasionally be seen at Mace Head. Consistently, a comparison shows regional European emissions derived from NAME at Mace Head (1999–2001) to be considerably smaller than those from the EXPORT campaign in 2000 (Supplementary Table 3). Furthermore, Jungfraujoch is often influenced by air masses from heavily industrial-

ized Northern Italy^{17,18}, but the excess concentrations of CH_3CCl_3 from this region are not significantly different from the rest of Europe. Therefore, high emissions attributed to Italy⁶ would have to originate specifically from southern Italy.

The reasons for the significant difference between our observation-based estimates and those from the EXPORT campaign are not obvious. Data from the EXPORT campaign are restricted to four days in summer 2000 (compared to the long-term measurements at our sites) and are more likely to be influenced by regional events of limited duration. Therefore, apart from any measurement and modelling uncertainties, we surmise that exceptional events occurred over Europe in summer 2000, which are not representative of longer time periods. Furthermore, analysed air masses in EXPORT could have been influenced by sources from Europe or other regions too remote to be detected from our sites.

We evaluated an upper-limit error due to the potentially inhomogeneous distribution of emissions, by separating pre-phase-out (1993) European (EU-15) CH_3CCl_3 emissions¹⁹ into those from regions significantly either influencing or not influencing our sites. For the 30% of emissions which historically would not have been reliably detected at our sites, a deliberately large positive error of 100% is added, which would at worst lead to estimated emissions of 0.9–5.6 Gg yr^{-1} in 2000–03 for Jungfraujoch, which is the site with the higher estimates.

Furthermore, to examine the consequences of making extrapolations from an unusually polluted subset of our data, we estimated emissions using inter-species correlation with only obvious elevated CH_3CCl_3 concentrations being used to detect pollution events. In this approach, point sources of limited duration (such as accidental release) could lead to significant overestimation of annual emis-

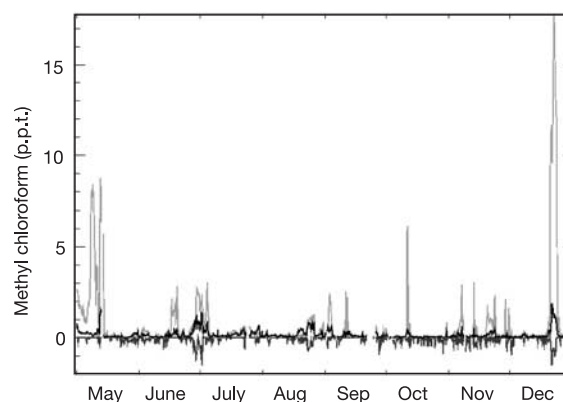


Figure 2 CH_3CCl_3 deviations from the baseline at Mace Head in 2000. Measured time series are shown as deviations (mirrored) below the zero line for clarity. Modelled time series using NAME-inversion European emission estimates of 0.9 Gg are shown in black and using NAME with EXPORT-derived emission estimates of 23 Gg (ref. 6) are shown in light grey.

sions. Resulting European CH_3CCl_3 emission estimates would be $5.0\text{--}7.3\text{ Ggyr}^{-1}$ for Jungfraujoch and $4.6\text{--}5.0\text{ Ggyr}^{-1}$ for Mace Head. As expected, figures from both this test and the inhomogeneity test exceed our best estimates given earlier, but are still much less than the EXPORT-based emission estimates.

Enhanced CH_3CCl_3 concentrations were also found in the MINOS campaign (Crete, 2001)⁶. From our sites it is not possible to infer emissions from this part of Europe. However, consumption-based estimates indicate minor emissions in Turkey at the end of the 1990s¹ and combined estimates for Turkey, Iran, Algeria, Egypt, Saudi Arabia, Jordan and Bahrain indicate ongoing emissions of 1.6, 1.3, 0.9 and 4.4 Ggyr^{-1} between 1999 and 2002 (refs 5, 20).

In conclusion, our long-term estimates of recent European emissions of CH_3CCl_3 are considerably lower than those from the EXPORT campaign⁶, but are higher than recent consumption-based figures¹. Our European emission estimates in the late 1990s and early 2000s indicate small-scale releases subsequent to the Montreal Protocol restrictions. These European emissions are, on their own, not enough to provide the $65\text{--}117\text{ Gg}$ additional global emissions of CH_3CCl_3 required to conclude a stable global OH radical abundance in recent years^{8,12}. Therefore, emissions of this magnitude, if real, would have to originate from other regions of the globe, and would indicate a severe under-reporting of emissions by industry and of consumption by the Montreal Protocol reporting process. □

Methods

Measurement sites and analytical methods

Mace Head is located on the west coast of Ireland and is part of AGAGE (Advanced Global Atmospheric Gases Experiment). Since 1994, high-frequency (40-min interval) measurements of the principal halocarbons and radiatively active trace gases have been made there by gas chromatography-electron capture detection (GC-ECD)².

Jungfraujoch (3,580 m above sea level) is situated in the northern part of the Swiss Alps in central Europe and is part of SOGE (System for Observation of Halogenated Greenhouse Gases in Europe). The station is influenced by air masses from the free troposphere and from the polluted European boundary layer. Since January 2000, 23 halocarbons have been analysed every four hours by gas chromatography-mass spectrometry (GC-MS)^{2,17}.

At both sites, each air measurement is bracketed with calibrations using real-air working standards. These standards were prepared by compressing background ambient air in 35-litre electropolished canisters (Essex) with a cleaned compressor (RIX, SA-3). Working standards were calibrated against real-air secondary standards compressed at Trinidad Head California and calibrated against primary standards on the SIO-98 calibration scale (Scripps Institution of Oceanography²). Checks for linearity and for blanks are performed regularly.

The inter-species correlation for emission estimation

For the determination of the excess concentrations during pollution events, the calculation of an accurate baseline is crucial. In method a, the baseline concentrations for CH_3CCl_3 and CO were iteratively identified for a 121-day period centred on each observation by a gaussian distribution of baseline values and removing observations exceeding the median plus three standard deviations. In method b, baseline concentrations were calculated iteratively, as the local linear regression of the concentrations. The time window was 90 days and two standard deviations were applied (except for CO in Mace Head, where four standard deviations were used). The standard deviation was calculated from the bracketing working real-air standards (that is, air measurements without pollution).

Both methods then used an identical scheme to produce emission estimates. Baselines were subtracted and European pollution events were identified by all excess events for CO and CH_3CCl_3 . At Mace Head this algorithm has already been proved equally effective in resolving short-term pollution events, as trajectory-based and carbon-dioxide-based approaches^{21,22}.

The yearly European CH_3CCl_3 emission estimates were then calculated by dividing the annual sum of the excess CH_3CCl_3 concentration by the annual sum of the excess CO concentration. Thereby, only data were included where concurrent measurements of both substances existed. These yearly ratios of excess concentrations were then multiplied by the CO emission estimates for Europe (EU-15 and Switzerland)¹⁶ (Supplementary Table 4). We thereby assumed that emissions from EU-15 regions that only slightly influence air masses at our two sites (Portugal, Spain, Southern Italy and Greece), do have similar ratios between CH_3CCl_3 and CO emissions, as in the regions actually sampled.

Lagrangian dispersion model NAME

The NAME model was used to generate maps of the origin of the air (within ten days of transport) arriving at Mace Head. NAME is driven by the three-dimensional synoptic meteorology from the UK Meteorological Office's numerical weather-prediction model. The fraction of air arriving from each point of origin at a grid volume $0.833^\circ \times 0.555^\circ \times 0\text{--}100\text{ m}$ centred on Mace Head is recorded for each one-hour period or each six-hour period for the one-yearly and the three-yearly estimates, respectively. The model domain covers $27^\circ\text{W--}44^\circ$

E and $25^\circ\text{N--}74^\circ\text{N}$ and calculates each air–origin map by following 100,000 air parcels (one-yearly estimate) or 60,000 air parcels (three-yearly estimates) backwards in time.

If the air has come mainly from the Atlantic Ocean, excluding southerly latitudes, and the local atmospheric conditions at Mace Head are well-mixed, then observations recorded during that period are considered to be the baseline. For the one-yearly (three-yearly) estimates excess concentrations were excluded before fitting a quadratic function to a rolling one-month (three-month) window of baseline points and then mean-smoothing within a rolling twenty-day (four-week) period. The result is an estimate of the time-series of mid-latitude Northern Hemisphere CH_3CCl_3 baseline concentrations.

The deviations from baseline were averaged over each one- or six-hour period with all negative mean values considered to be zero. This time series was linked with the NAME model output to estimate the emission distribution of CH_3CCl_3 over western Europe. The iterative best-fit technique (simulated annealing) is used to derive these values based on a statistical score, combining correlation coefficient, normalized mean square error and fraction within a factor of two¹⁵. Three-year rolling averages of the emissions of CH_3CCl_3 over several geographical areas have been estimated using this approach.

The ability of the NAME model to estimate European emissions was tested against CO emissions provided by EMEP¹⁶ and the inferred European source strength was well in accordance with this data (Supplementary Fig. 1).

Received 20 January; accepted 16 November 2004; doi:10.1038/nature03220.

- McCulloch, A. & Midgley, P. M. The history of methyl chloroform emissions: 1951–2000. *Atmos. Environ.* **35**, 5311–5319 (2001).
- Prinn, R. G. *et al.* A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE. *J. Geophys. Res.—Atmos.* **105**, 17751–17792 (2000).
- Montzka, S. A. *et al.* Present and future trends in the atmospheric burden of ozone-depleting halogens. *Nature* **398**, 690–694 (1999).
- Montzka, S. A. & Fraser, P. J. *Scientific Assessment of Ozone Depletion 2002* Ch. 1.1–1.81 (WMO Global Ozone Research and Monitoring Project—Report No. 47, Geneva, 2003).
- UNEP Production and Consumption of Ozone Depleting Substances under the Montreal Protocol, 1986–2000. (United Nations Environment Programme, Ozone Secretariat, Nairobi, 2002).
- Krol, M. C. *et al.* Continuing emissions of methyl chloroform from Europe. *Nature* **421**, 131–135 (2003).
- Prinn, R. G. *et al.* Atmospheric trends and lifetime of CH_3CCl_3 and global OH concentrations. *Science* **269**, 187–192 (1995).
- Prinn, R. G. *et al.* Evidence for substantial variations of atmospheric hydroxyl radicals in the past two decades. *Science* **292**, 1882–1888 (2001).
- Montzka, S. A. *et al.* New observational constraints for atmospheric hydroxyl on global and hemispheric scales. *Science* **288**, 500–503 (2000).
- Spivakovskiy, C. M. *et al.* Three-dimensional climatological distribution of tropospheric OH: Update and evaluation. *J. Geophys. Res.—Atmos.* **105**, 8931–8980 (2000).
- Krol, M., van Leeuwen, P. J. & Lelieveld, J. Global OH trend inferred from methylchloroform measurements. *J. Geophys. Res.—Atmos.* **103**, 10697–10711 (1998).
- Krol, M. & Lelieveld, J. Can the variability in tropospheric OH be deduced from measurements of 1,1,1-trichloroethane (methyl chloroform)? *J. Geophys. Res.—Atmos.* **108**, 4125 (2003).
- Gros, V. *et al.* Origin of anthropogenic hydrocarbons and halocarbons measured in the summertime European outflow (on Crete in 2001). *Atmos. Chem. Phys.* **3**, 1223–1235 (2003).
- Ryall, D. B., Derwent, R. G., Manning, A. J., Simmonds, P. G. & O'Doherty, S. Estimating source regions of European emissions of trace gases from observations at Mace Head. *Atmos. Environ.* **35**, 2507–2523 (2001).
- Manning, A. J., Ryall, D. B., Derwent, R. G., Simmonds, P. G. & O'Doherty, S. Estimating European emissions of ozone-depleting and greenhouse gases using observations and a modeling back-attribution technique. *J. Geophys. Res.—Atmos.* **108**, 4405 (2003).
- Vestreng, V. *et al.* Inventory Review 2004, Emission Data reported to CLRTAP and under the NEC Directive. (EMEP/EEA Joint Review Report, EMEP/MSC-W Note 1/2004, 2004).
- Reimann, S. *et al.* Halogenated greenhouse gases at the Swiss High Alpine Site of Jungfraujoch (3580 m asl): Continuous measurements and their use for regional European source allocation. *J. Geophys. Res.—Atmos.* **109**, D05307, doi:10.1029/2003JD003923 (2004).
- Forrer, J. *et al.* Variability of trace gases at the high-Alpine site Jungfraujoch caused by meteorological transport processes. *J. Geophys. Res.—Atmos.* **105**, 12241–12251 (2000).
- Midgley, P. M. & McCulloch, A. The production and global distribution of emissions to the atmosphere of 1,1,1-trichloroethane (methyl chloroform). *Atmos. Environ.* **29**, 1601–1608 (1995).
- UNEP Report of the Fifteenth Meeting of the Parties to the Montreal Protocol on Substances that Deplete the Ozone Layer (United Nations Environment Programme, 2003).
- O'Doherty, S. *et al.* In situ chloroform measurements at Advanced Global Atmospheric Gases Experiment atmospheric research stations from 1994 to 1998. *J. Geophys. Res.—Atmos.* **106**, 20429–20444 (2001).
- Cunnold, D. M. *et al.* In situ measurements of atmospheric methane at GAGE/AGAGE sites during 1985–2000 and resulting source inferences. *J. Geophys. Res.—Atmos.* **107**, 4225 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank C. Harth and P. Salameh for calibration and data processing, G. Spain and D. Brown for instrument operation and maintenance at Mace Head and the Physics Department of NUI Galway for access to Mace Head facilities. Measurements and modelling at Mace Head are supported by NASA (USA) and DEFRA (UK). We thank SAEFL/BUWAL and BBW for financial support, HFSJG for access to Jungfraujoch facilities and P. Kaufmann (MeteoSwiss) for meteorological data. Both sites receive financial support from the EU 5th Framework Program (project SOGE).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.R. (stefan.reimann@empa.ch).

Global azimuthal seismic anisotropy and the unique plate-motion deformation of Australia

Eric Debayle¹, Brian Kennett² & Keith Priestley³

¹*Institut de Physique du Globe de Strasbourg, Ecole et Observatoire des Sciences de la Terre, Centre National de la Recherche Scientifique and Université Louis Pasteur, 61084 Strasbourg, Cedex, France*

²*Research School of Earth Sciences, The Australian National University, Canberra ACT 0200, Australia*

³*Bullard Laboratories, University of Cambridge, Cambridge CB3 0EZ, UK*

Differences in the thickness of the high-velocity lid underlying continents as imaged by seismic tomography, have fuelled a long debate on the origin of the 'roots' of continents^{1–5}. Some of these differences may be reconciled by observations of radial aniso-

tropy between 250 and 300 km depth, with horizontally polarized shear waves travelling faster than vertically polarized ones². This azimuthally averaged anisotropy could arise from present-day deformation at the base of the plate, as has been found for shallower depths beneath ocean basins⁶. Such deformation would also produce significant azimuthal variation, owing to the preferred alignment of highly anisotropic minerals⁷. Here we report global observations of surface-wave azimuthal anisotropy, which indicate that only the continental portion of the Australian plate displays significant azimuthal anisotropy and strong correlation with present-day plate motion in the depth range 175–300 km. Beneath other continents, azimuthal anisotropy is only weakly correlated with plate motion and its depth location is similar to that found beneath oceans. We infer that the fast-moving Australian plate contains the only continental region with a sufficiently large deformation at its base to be transformed into azimuthal anisotropy. Simple shear leading to anisotropy with a plunging axis of symmetry may explain the smaller azimuthal anisotropy beneath other continents.

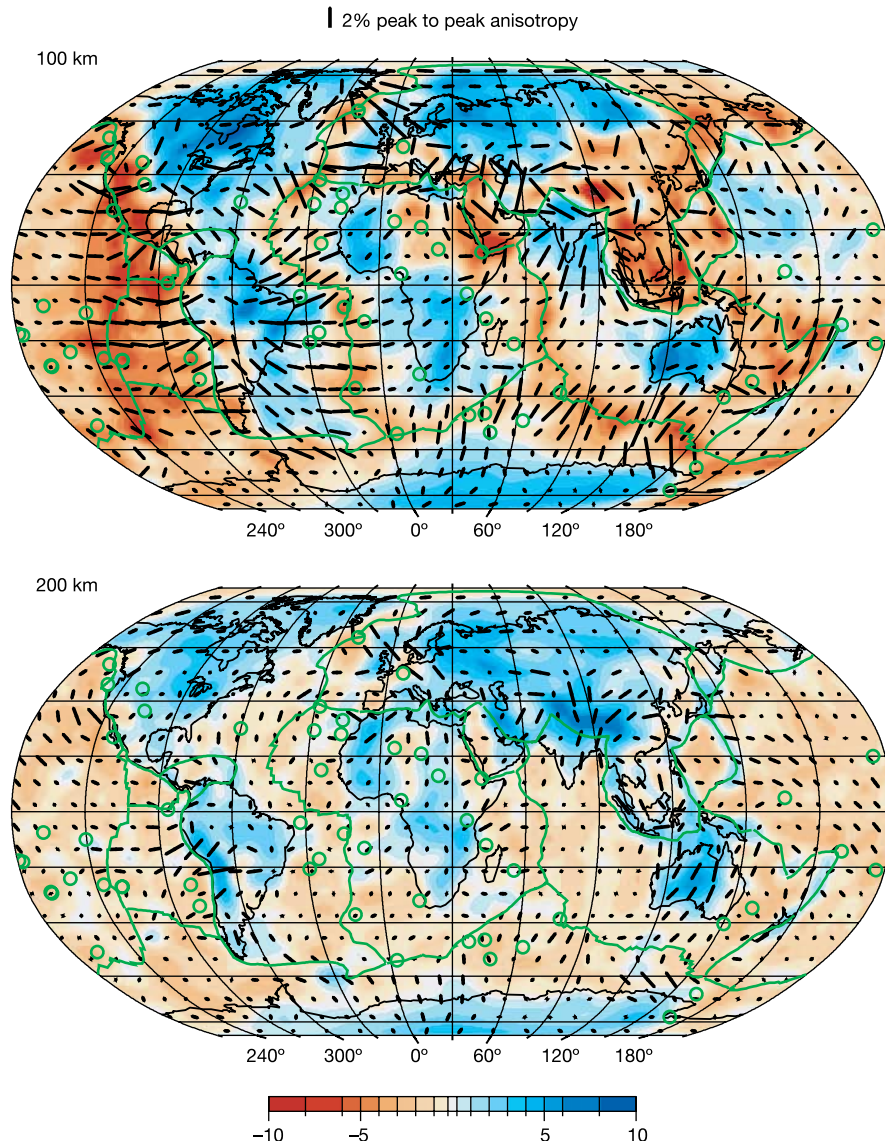


Figure 1 SV-wave heterogeneity and azimuthal anisotropy (black bars oriented along the axis of fast propagation) at 100 and 200 km depth obtained from the inversion of 100,779 Rayleigh waveforms. Hotspot locations are indicated by green circles. The length of the

black bars is proportional to the maximum amplitude of azimuthal anisotropy (bar length for 2% peak to peak anisotropy shown at top). SV-wave perturbations (in per cent relative to PREM) are represented with the colour scale.

Our finding comes in the context of the building of a new surface wave tomographic model of the upper mantle designed to investigate shear wave heterogeneities and azimuthal anisotropy. In the upper mantle, horizontally travelling surface waves provide better vertical resolution than body waves, which are more subject to vertical smearing owing to their steep incidence. The novelty of our approach is to directly extract from surface waves the directions of fast propagation for horizontally travelling 'SV' (vertically polarized) waves, with an unprecedented lateral resolution at a global scale, instead of determining anisotropy in group or phase velocity, as in other recent studies^{8,9}. Direct estimates of fast SV directions provide better vertical resolution compared to group or phase velocity measurements, which represent a weighted average of the structure/anisotropy over a frequency-dependent depth interval. Our path coverage (see Supplementary Fig. 1) allows us to resolve anisotropic variations with horizontal wavelengths matching the scale of the lithospheric blocks that have coalesced to form continents (about 1,000–1,500 km). We can therefore better differentiate between different ways of producing anisotropy: the effect of plate motion would be expected to be smooth at continental scale, whereas ancient deformation frozen in the lithosphere displays shorter-scale lateral variations owing to the complex tectonic history of continents.

There is no systematic difference in the depth location of azimuthal anisotropy between continents and oceans, except beneath Australia. Figure 1 shows the azimuthal variations for SV waves superimposed on the pattern of seismic heterogeneities at 100 and 200 km depth in the upper mantle. At 100 km depth, the amplitude of azimuthal anisotropy generally exceeds 2% beneath young oceans, and exhibits highly variable amplitudes beneath continents and old oceanic basins. When old continents are underlain by significant azimuthal anisotropy (as in Australia, India, and North and South America), the anisotropy varies over short horizontal distances. At 200 km depth, the only continent where plate-scale anisotropy is larger than 2% is Australia. Other continents are associated with weak anisotropy, except locally beneath the Tibetan plateau and the Andean subduction zone. Plate-scale anisotropy has also disappeared beneath oceans except beneath the northern Pacific, where weak (<1%) azimuthal anisotropy extends over a broad region.

Figure 2a shows the average amplitude of azimuthal anisotropy as a function of depth in the upper mantle, calculated for Australia, other continents, and oceanic basins. In oceanic regions, the maximum amplitude of azimuthal anisotropy occurs near 100 km depth, in agreement with previous studies that have suggested an intense deformation at the base of oceanic plates^{3,6}. The depth configuration of azimuthal anisotropy beneath Australia in the depth range 150–300 km is similar to that beneath oceans in the depth range 50–200 km but shifted downward by 100 km. Except in the upper 100 km, Australia is completely different from other continents, which display a gentle decrease of anisotropy from 1.4% at 50 km to about 0.6% at 300 km depth.

Further, Australia is the only continental plate where azimuthal anisotropy correlates significantly with present-day plate motion. This peculiar behaviour of the Australian continent is highlighted in the global correlation (Fig. 3) between azimuthal anisotropy and the present-day absolute plate motion (APM). Fast anisotropy directions beneath Australia do not correlate with APM at depths shallower than 150 km, but show strong correlation from 150 km to 300 km depth with a maximum near 200 km. This agrees with previous regional surface wave tomography for the continent^{10–13}. None of the other continents show significant plate-scale correlation between anisotropic directions and APM. The anisotropic signature of the Australian plate is similar to that observed at shallower depths beneath oceans. In young oceanic regions where the lithosphere is expected to be thin, significant correlation

between anisotropy and APM is observed at 50 km depth (Fig. 3) and extends over large regions around the mid-ocean ridges beneath the Pacific, Indian and Atlantic Oceans. At 100 and 150 km depth, the regions where anisotropy correlates with APM shift to the old oceanic basins. This shift suggests that the depth of plate-motion-induced deformation increases with the age of the sea floor and the thickness of the oceanic lithosphere. Figure 2b shows the average correlation between APM and fast anisotropic directions calculated for Australia, other continents, and oceanic basins. The strong correlation between Australian anisotropy and APM is prominent between 150 and 300 km depth. The fast direction beneath the oceans displays, on average, a weaker correlation with APM between 100 and 250 km depth. This weaker correlation between oceanic anisotropy and APM can be related to the observation⁹ that azimuthal anisotropy beneath oceans aligns better with the largest axis of the finite-strain ellipsoid than with the absolute plate motion. Azimuthal anisotropy of continents other than Australia does not correlate at any depth with APM.

Although an accurate prediction of SKS splitting from surface wave azimuthal anisotropy models is not possible (see Methods), we believe that our results provide a qualitative explanation for the

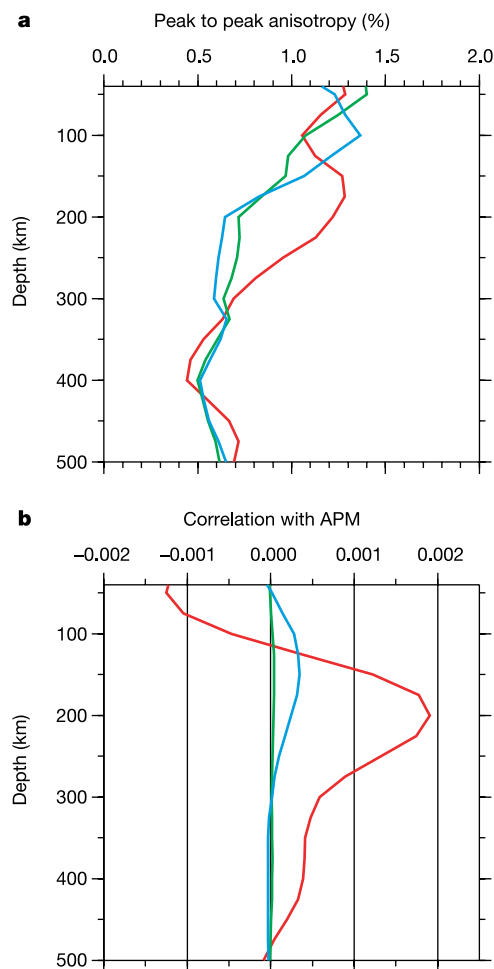


Figure 2 Azimuthal anisotropy amplitude and correlation with plate motion for different tectonic provinces. Green, continents except Australia; red, Australia; and blue, oceans. **a**, Amplitude of peak to peak azimuthal anisotropy as a function of depth. **b**, Correlation between the fast direction of SV waves and absolute plate motion as a function of depth. The correlation coefficient is computed as in Fig. 3, but averaged over the different tectonic provinces.

differences in SKS observations between Australia and the other continents¹⁴. SKS studies generally show typical delay times close to 1 s in most continental regions¹⁵, but only null¹⁶ or very weak¹⁷ splitting beneath Australia. A vertically travelling wave passing through two anisotropic layers with orthogonal directions as found beneath Australia (Fig. 1) would undergo a null or very weak splitting if each layer produced a similar time separation between the fast and slow polarized S-waves. Beneath other continents, weak influence of basal drag on the lithosphere may explain why azimuthal anisotropy is observed only in a layer located in the uppermost 100 km of the mantle. The complex organization of surface wave azimuthal anisotropy present within this layer, and the good agreement between SKS fast directions and fossil geological trends generally observed for continents other than Australia, form the basis for attributing this shallow anisotropic layer (whose thickness is compatible with typical observed SKS delay times) to deformation frozen in the lithosphere. The similar depth behaviour and complex organization of anisotropy down to about 150 km depth observed beneath Australia and other continents (Figs 1 and 2a) suggests that frozen deformation in the lithosphere is also the explanation of shallow upper mantle azimuthal anisotropy beneath Australia.

Laboratory experiments on olivine aggregates suggest that simple shear at the base of a moving plate will produce anisotropy in olivine with a fast a axis that follows the principal extension direction for modestly deformed aggregates and aligns with the direction of flow for large deformation¹⁸. Complications occur under water-rich conditions¹⁹ that are not common in the upper mantle²⁰ and are probably confined to subduction zones or regions where upwelling material contains a large amount of water¹⁹. For relatively water-poor olivine, modest simple shear should therefore produce anisotropy with a plunging a axis. For surface waves, although the azimuthal variation of Rayleigh wave velocity gradually reduces when the a axis departs from the horizontal, the direction of the fastest Rayleigh waves remains in the vertical plane containing the a

axis²¹. Our results therefore suggest that the Australian plate is the only continental plate whose motion is fast enough to produce large scale deformation at its base. The slower horizontal motion of other continental plates may produce smaller basal deformation, and thus a larger proportion of olivine crystals with a plunging axis of symmetry and a weaker azimuthal anisotropy. At depths greater than 220 km, enrichment in clinopyroxene may also contribute to the vanishing of anisotropy²².

This basal drag mechanism can explain both the radial and the azimuthal anisotropy of the Australian continent. If the lattice-preferred orientation of olivine crystals is the main mechanism responsible for upper mantle anisotropy⁷, the existence of significant azimuthal anisotropy with coherent directions over broad regions beneath Australia implies that radial anisotropy with SH (horizontally polarized) waves faster than SV waves should also be present to the same depth¹². Radial anisotropy with SH velocities greater than SV velocities ('SH > SV') has indeed been observed beneath Australia down to at least 250 km depth in previous regional^{12,23} and global studies².

Weak azimuthal anisotropy as observed under other continents is compatible with a modest SH > SV radial anisotropy depending on the dip and proportion of oriented olivine crystals. However, the large SH > SV radial anisotropy observed at the base of continents from 250 to 400 km depth by Gung *et al.*² is hard to reconcile with our observation of small azimuthal anisotropy and with the typical 1 s delay time in SKS studies. This incompatibility provides a global scale illustration of the problem of explaining the amplitude of surface wave radial anisotropy with current petrological models, a well-documented problem in regional studies^{3,12,24}. To achieve large SH > SV radial anisotropy with weak azimuthal anisotropy, the olivine crystals need to be preferentially aligned in the horizontal plane, but randomly oriented. Small scale convection due to irregularities of the base of the high-velocity lid or perturbations by mantle plumes can produce such effects.

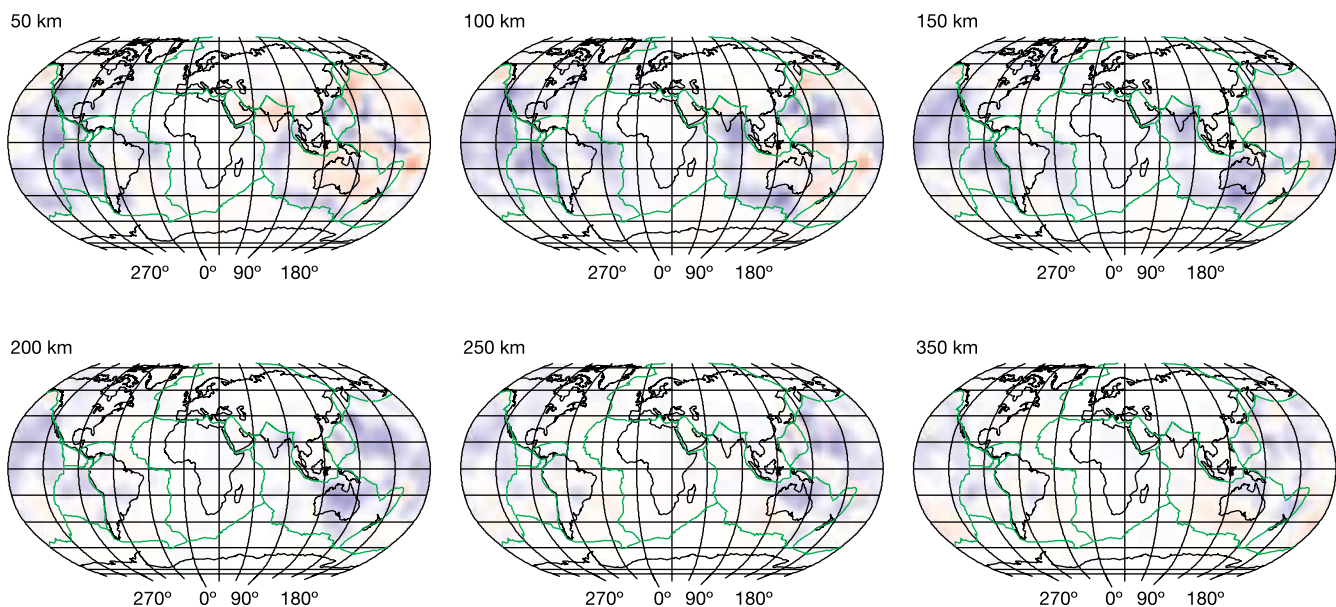


Figure 3 Correlation between fast direction of SV waves and plate motion directions. APM is derived from model Nuvel-1³¹ by imposing a null average rotation of the lithosphere. The correlation coefficient is defined as $|\text{FastSV}||\text{APM}| \cos[2(\phi)]$, where FastSV is the fast SV wave vector, APM is the absolute plate motion vector and ϕ is the angle between the two vectors. Good correlation (parallelism of the two vectors) is represented in blue, weak correlation is represented in white and bad correlation

(orthogonality) is represented in red. The colour scales are symmetric, adapted to cover the full range of values at each depth (shown at top left of each panel). The amplitude of the correlation coefficient is not shown, as it depends on the strength of anisotropy and is difficult to interpret. Plate boundaries are indicated with green lines.

Our model provides an explanation for SKS observations in continental regions, on the simple basis that anisotropy due to plate motion differs between continental plates. Additional data will be needed to investigate in more detail the global pattern of radial anisotropy in seismic parameters and to reconcile it with other anisotropic observations. Although seismic anisotropy provides a unique way to investigate deformation of the upper mantle, it is increasingly clear that the assumptions underlying anisotropic observations must continue to be questioned. □

Methods

Our model is an 'SV' model constrained by fundamental and higher mode Rayleigh waves (as in ref. 4), but we include azimuthal anisotropy in the inversion and include many short epicentre–station paths (see Supplementary Fig. 2) to improve the lateral resolution of upper mantle structure²⁵. The data set consists of 100,779 Rayleigh waveforms that provide a dense global coverage (see Supplementary Fig. 1), although variations due to the uneven distribution of events and station are inevitable. The number of crossing rays per cell (400×400 km) varies between 20 and 2,800 for continents, ensuring redundancy in the data even where ray coverage is the poorest (Africa and Antarctica).

Model construction

Our model is built using a two step tomographic procedure²⁵. First, an automated nonlinear waveform inversion technique^{10,26} is used to model each individual Rayleigh waveform in terms of a depth-dependent SV-wave velocity model representing the average mantle structure along the path. The one-dimensional path-average models are then combined in a tomographic inversion²⁵ to retrieve simultaneously the three-dimensional SV-wave velocity structure and the azimuthal anisotropy of SV waves. The procedure exploits the azimuthal dependence of the Rayleigh wave phase and group velocities in a slightly anisotropic medium²⁷. This azimuthal dependence contains terms in $\cos(2\theta)$, $\sin(2\theta)$ and $\cos(4\theta)$, $\sin(4\theta)$, where θ is the azimuth relative to north, and has been recently inverted to retrieve the global azimuthal variations in the group or phase velocities at different periods^{8,9}. The azimuthal terms can also be inverted in depth, as they depend on several combinations of the elastic parameters via a set of partial derivatives proportional to the partial derivatives of a transversely isotropic medium with a vertical axis of symmetry²⁸. The combinations of elastic parameters best resolved by Rayleigh waves involve at each depth an isotropic term that corresponds to the SV-wave velocity and two azimuthal terms that display a variation in $\cos(2\theta)$, $\sin(2\theta)$ relative to the direction of maximum velocity. In a long period approximation, the one-dimensional SV-wave velocity models obtained after the waveform fitting depend on these three combinations of elastic parameters, which control the velocities of SV waves propagating horizontally for azimuth θ (ref. 29). For an olivine model, the direction of maximum velocity coincides with the horizontal projection of the fast a axis of olivine crystals. Both the non-azimuthal term and the direction of fast seismic velocities extracted from the $\cos(2\theta)$, $\sin(2\theta)$ azimuthal terms are represented on Fig. 1.

Resolution issues and model tests

A Voronoi diagram built using the approach of ref. 25 (see Supplementary Fig. 3) provides a useful guide to our ability to resolve azimuthal anisotropy with our data coverage. The cells in Supplementary Fig. 3 are the smallest for which the local distribution of rays ensures that the $\cos(2\theta)$, $\sin(2\theta)$ SV-wave azimuthal variation can be resolved. This 'optimized Voronoi' diagram is based on the ray distribution alone and does not incorporate any *a priori* information on the data or the model, nor does it take into account the influence of different parameter choices. However, it provides a useful proxy for resolution when combined with the horizontal degree of smoothing imposed *a priori* on the tomographic inversion. We impose lateral smoothness through a gaussian correlation function with a standard deviation of 400 km, chosen to minimize the trade-off between isotropic and anisotropic parameters. Azimuthal anisotropy is guaranteed to be resolved when the size of the Voronoi cells in Supplementary Fig. 3 is smaller than, or comparable to, a circular surface with a minimum diameter of about 1,000 km. Azimuthal anisotropy resolution is achieved for 1,000-km circular regions beneath most continents and at the Voronoi level for western Africa and Antarctica.

We have performed a variety of tests using real and synthetic data to estimate potential leakage by non-inverted parameters, vertical resolution and how well our model can predict SKS observations. We have found (1) that contamination by non-inverted parameters, such as the 4θ azimuthal variation of the Rayleigh waves phase velocity or the B and H combinations of elastic coefficients as defined in ref. 28, is weak (Supplementary Fig. 4), (2) that our vertical resolution is sufficient to recover a change in anisotropic direction at the bottom of continental plates (Supplementary Fig. 5), and (3) that it is not possible to make accurate predictions of SKS observations from our tomographic model. In fact, the length scales over which SKS results vary in continental regions are often below the horizontal resolution of surface waves. Furthermore, vertical smearing (Supplementary Fig. 5) can bias the splitting prediction (Supplementary Fig. 6b), and the SKS predictions themselves (Supplementary Fig. 6c) rely on a theory³⁰ that does not properly handle inclined symmetry axes and is only effective for rather long period waves.

Received 2 July; accepted 29 November 2004; doi:10.1038/nature03247.

1. Jordan, T. H. The continental tectosphere. *Rev. Geophys.* **13**, 1–12 (1975).

- Gung, Y., Panning, M. & Romanowicz, B. Global anisotropy and the thickness of continents. *Nature* **422**, 707–711 (2003).
- Ekström, G. & Dziewonski, A. M. The unique anisotropy of the Pacific upper mantle. *Nature* **394**, 168–172 (1998).
- Ritsema, J., van Heijst, H. J. & Woodhouse, J. H. Global transition zone tomography. *J. Geophys. Res.* **109**, B02302, doi:10.1029/2003JB002610 (2004).
- Montagner, J. P. & Tanimoto, T. Global upper mantle tomography of seismic velocities and anisotropies. *J. Geophys. Res.* **96**, 20337–20351 (1991).
- Forsyth, D. W. The early structural evolution and anisotropy of the oceanic upper mantle. *Geophys. J. R. Astron. Soc.* **43**, 103–162 (1975).
- Nicolas, A. & Christensen, N. I. in *Composition, Structure and Dynamics of the Lithosphere-Asthenosphere System* (eds Fuchs, F. & Froidevaux, C.) 111–123 (Geodyn. Ser. Vol. 16, AGU, Washington DC, 1987).
- Trampert, J. & Woodhouse, J. H. Global anisotropic phase velocity maps for the fundamental mode surface waves between 40 and 150 s. *Geophys. J. Int.* **154**, 154–165 (2003).
- Becker, T. W., Kellogg, J. B., Ekström, G. & O'Connell, R. J. Comparison of azimuthal seismic anisotropy from surface waves and finite strain from global mantle-circulation models. *Geophys. J. Int.* **155**, 696–714 (2003).
- Debayle, E. SV-wave azimuthal anisotropy in the Australian upper-mantle: Preliminary results from automated Rayleigh waveform inversion. *Geophys. J. Int.* **137**, 747–754 (1999).
- Debayle, E. & Kennett, B. L. N. The Australian continental upper mantle: structure and deformation inferred from surface waves. *J. Geophys. Res.* **105**, 25423–25450 (2000).
- Debayle, E. & Kennett, B. L. N. Anisotropy in the Australasian upper mantle from Love and Rayleigh waveform inversion. *Earth Planet. Sci. Lett.* **184**, 339–351 (2000).
- Simons, F. J. & van der Hilst, R. D. Seismic and mechanical anisotropy and the past and present deformation of the Australian lithosphere. *Earth Planet. Sci. Lett.* **211**, 271–286 (2003).
- Park, J. & Levin, V. Seismic anisotropy: Tracing plate dynamics in the mantle. *Science* **296**, 485–489 (2002).
- Savage, M. K. Seismic anisotropy and mantle deformation: What have we learned from shear wave splitting? *Rev. Geophys.* **37**, 65–106 (1999).
- Ozalabey, S. & Chen, W. P. Frequency dependent analysis of SKS/SKKS waveforms observed in Australia: evidence for null birefringence. *Phys. Earth Planet. Inter.* **114**, 197–210 (1999).
- Clitheroe, G. & van der Hilst, R. D. in *Structure and Evolution of the Australian Continent* (eds Braun, J., Dooley, J., Goleby, B., van der Hilst, R. & Klootwijk, C.) 73–78 (Geodyn. Monogr. 26, AGU, Washington DC, 1998).
- Zhang, S. & Karato, S. I. Lattice preferred orientation of olivine aggregate deformed in simple shear. *Nature* **375**, 774–777 (1995).
- Jung, H. & Karato, S. I. Water-induced fabric transitions in olivine. *Science* **293**, 1460–1463 (2001).
- Kaminski, E. The influence of water on the development of lattice preferred orientation in olivine aggregates. *Geophys. Res. Lett.* **29**, 12, doi:10.1029/2002GL014710 (2002).
- Babuska, V. & Cara, M. *Seismic Anisotropy in the Earth* (Kluwer Academic, Dordrecht, 1991).
- Beghein, C. & Trampert, J. Probability density functions for radial anisotropy: implications for the upper 1200 km of the mantle. *Earth Planet. Sci. Lett.* **217**, 151–162 (2003).
- Gaherty, J. B. & Jordan, T. Lehmann discontinuity as the base of an anisotropic layer beneath continents. *Science* **268**, 1468–1471 (1995).
- Maupin, V. & Cara, M. Love-Rayleigh wave incompatibility and possible deep upper mantle anisotropy in the Iberian peninsula. *Pure Appl. Geophys.* **138**, 429–444 (1992).
- Debayle, E. & Sambridge, M. Inversion of massive surface wave data sets: Model construction and resolution assessment. *J. Geophys. Res.* **109**, B02316 doi:10.1029/2003JB002652, (2004).
- Cara, M. & Lévêque, J. J. Waveform inversion using secondary observables. *Geophys. Res. Lett.* **14**, 1046–1049 (1987).
- Smith, M. & Dahlen, F. The azimuthal dependence of Love and Rayleigh wave propagation in a slightly anisotropic medium. *J. Geophys. Res.* **78**, 3321–3333 (1973).
- Montagner, J. P. & Nataf, H. C. A simple method for inverting the azimuthal anisotropy of surface waves. *J. Geophys. Res.* **91**, 511–520 (1986).
- Lévêque, J. J., Debayle, E. & Maupin, V. Anisotropy in the Indian Ocean upper mantle from Rayleigh- and Love-waveform inversion. *Geophys. J. Int.* **133**, 529–540 (1998).
- Montagner, J. P., Griot-Pommerehne, D. A. & Lavé, J. How to relate body wave and surface wave anisotropy? *J. Geophys. Res.* **105** (B8), 19015–19027 (2000).
- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Current plate motion. *Geophys. J. Int.* **101**, 425–478 (1990).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by programme DyETI conducted by the French Institut National des Sciences de l'Univers (INSU). The data used in this work were obtained from the GEOSCOPE, GDSN, IDA, MEDNET and GTSN permanent seismograph networks, and completed with data collected after the PASSCAL broadband experiments, the SKIPPY and subsequent broadband deployments in Australia, and the INSU deployments in the Horn of Africa and the Pacific (PLUME experiment). Supercomputer facilities were provided by the IDRIS and CINES national centres in France. Special thanks to J. M. Brendle at EOST for technical support, S. Fishwick for providing broadband data from the Western Australian craton field deployment, the staff of the Research School of Earth Science who collected the SKIPPY data in the field, and A. Maggi for suggestions that improved the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.D. (Eric.Debayle@eost.u-strasbg.fr).

Effective leadership and decision-making in animal groups on the move

Iain D. Couzin^{1,2}, Jens Krause³, Nigel R. Franks⁴ & Simon A. Levin¹

¹Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA

²Department of Zoology, South Parks Road, University of Oxford, Oxford OX1 3PS, UK

³Centre for Biodiversity and Conservation, School of Biology, University of Leeds, Leeds LS2 9JT, UK

⁴Centre for Behavioural Biology, School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK

For animals that forage or travel in groups, making movement decisions often depends on social interactions among group members^{1,2}. However, in many cases, few individuals have pertinent information, such as knowledge about the location of a food source^{3,4}, or of a migration route^{5–9}. Using a simple model we show how information can be transferred within groups both without signalling and when group members do not know which individuals, if any, have information. We reveal that the larger the group the smaller the proportion of informed individuals needed to guide the group, and that only a very small proportion of informed individuals is required to achieve great accuracy. We also demonstrate how groups can make consensus decisions, even though informed individuals do not know whether they are in a majority or minority, how the quality of their information compares with that of others, or even whether there are any other informed individuals. Our model provides new insights into the mechanisms of effective leadership and decision-making in biological systems.

Primary questions concerning the mechanism of information transfer in groups include how uninformed individuals recognize those that are informed, whether such recognition is actually necessary, and how groups can come to a collective decision when informed individuals differ in preference¹⁰. It is known that several animal species have evolved specific recruitment signals that help guide conspecifics. Most famous in this context is the waggle-dance of the honeybee that recruits hive members to visit food sources^{5,7,8,11}. Furthermore, valuable experience may be correlated with age or dominance^{1,2}, which can presumably be estimated by conspecifics of some species¹². However, it remains questionable whether such explanations hold when migrating groups of fish, ungulates, insects and birds are considered, where crowding limits the range over which individuals can detect one another^{1,2}. In pelagic fish schools, for example, individuals are usually less than one body-length apart¹³. Although it is likely that some species have a genetically determined propensity to migrate in a general direction^{14,15}, or respond to abiotic cues such as thermal gradients that may aid migration^{16,17}, it is likely for many species that experienced group members play an important role in guiding those that are less experienced or inexperienced. Relatively few informed individuals within fish schools are known to be able to influence the foraging behaviour of the group³ and the ability of a school to navigate towards a target⁴. Similarly, very few individuals (approximately 5%) within honeybee swarms can guide the group to a new nest site⁷.

Furthermore, for some animal groups such as large insect swarms or fish schools, it may be unreasonable to assume that group members have the capacity for individual recognition. Here we address two fundamental issues, both occurring in the absence of complex signalling mechanisms and when it is not possible for group members to establish who has and has not got information. First, how information about the location of resources, or of a migration route, can be transferred within groups; and second, how

individuals can achieve a consensus when informed individuals differ in their preferences.

We take into account the ability of grouping individuals to modify their motion on the basis of that of local neighbours (social interactions)^{2,18,19}. Groups are composed of N individuals. Each individual with position vector $\mathbf{c}_i(t)$, direction vector $\mathbf{v}_i(t)$, and speed s_i , attempts to maintain a minimum distance α between itself i and others j at all times by turning away from neighbours within that range

$$\mathbf{d}_i(t + \Delta t) = - \sum_{j \neq i} \frac{\mathbf{c}_j(t) - \mathbf{c}_i(t)}{|\mathbf{c}_j(t) - \mathbf{c}_i(t)|} \quad (1)$$

where $\mathbf{d}_i$ represents a desired direction of travel. This simulates individuals acting to maintain personal space and to avoid collisions^{1,2}. Avoidance is the highest priority. If neighbours are not detected within this region then the individual will tend to become attracted towards, and aligned with^{1,2,11,13,18}, j neighbours within a local interaction range ρ :

$$\mathbf{d}_i(t + \Delta t) = \sum_{j \neq i} \frac{\mathbf{c}_j(t) - \mathbf{c}_i(t)}{|\mathbf{c}_j(t) - \mathbf{c}_i(t)|} + \sum_{j \neq i} \frac{\mathbf{v}_j(t)}{|\mathbf{v}_j(t)|} \quad (2)$$

Here $\mathbf{d}_i(t + \Delta t)$ is converted to the corresponding unit vector $\hat{\mathbf{d}}_i(t + \Delta t) = \mathbf{d}_i(t + \Delta t) / |\mathbf{d}_i(t + \Delta t)|$.

To incorporate the influence of informed group members, a proportion of the individuals p are given information about a preferred direction (simulated as a unit vector $\mathbf{g}$) representing, for example, the direction to a known resource, or a segment of a migration route. All other individuals are naive and have no preference to move in any particular direction, and are also not informed as to which individuals within the group have information and which do not. Informed individuals balance the influence of their preferred direction and their social interactions with weighting term ω , and replacing $\hat{\mathbf{d}}_i(t + \Delta t)$ by $\mathbf{d}_i'(t + \Delta t)$, where:

$$\mathbf{d}_i'(t + \Delta t) = \frac{\hat{\mathbf{d}}_i(t + \Delta t) + \omega \mathbf{g}_i}{|\hat{\mathbf{d}}_i(t + \Delta t) + \omega \mathbf{g}_i|} \quad (3)$$

If $\omega = 0$, vector $\mathbf{g}_i$ has no influence and individuals have no desire to move in any specific direction. As ω approaches 1, individuals tend to balance their preference to move in direction $\mathbf{g}_i$ with their desire to maintain social interactions with group members. As ω exceeds 1, individuals are more heavily influenced by their preferred direction $\mathbf{g}_i$ than by their neighbours (see Methods section for details of the model).

The accuracy of the group, and hence the quality of information transfer, can be quantified as the normalized angular deviation²⁰ of group direction (see Methods) around the preferred direction $\mathbf{g}$, with a minimum value of 0 and maximum of 1.

For a given group size we found that the accuracy of group motion (in a preferred direction) increased asymptotically as the proportion of informed individuals increased. Furthermore, as group size became larger this relationship became increasingly nonlinear (Fig. 1a), meaning that the larger the group, the smaller the proportion of informed individuals needed to guide the group with a given accuracy. Thus for sufficiently large groups only a very small proportion of informed individuals is needed to achieve close to maximal accuracy. For animal groups such as migrating honeybee colonies, there are likely to be costs associated with increasing the proportion of scouts (informed individuals) within the colony owing to the time taken to recruit others via the waggle-dance^{5–9,11}, and for scouts to learn essential navigational skills^{7,8}. Thus, we may expect colonies to have evolved such that they have effectively reached the asymptote of accuracy (Fig. 1a) and we predict that they would achieve little benefit in having larger scout populations.

The influence of the weighting of preferred direction ω was of least importance if the proportion of informed individuals is small

or large (Fig. 2a, c, d). At intermediate values, however, ω becomes strongly positively correlated with group accuracy (Fig. 2b). This corresponds to the region in Fig. 1 where the rate of increase in accuracy reaches its asymptote. There is a cost to increasing ω for these intermediate values of p , however, because there exists a trade-off between the accuracy of group motion and the probability that the group will fragment (Fig. 2b). Lack of group cohesion would clearly be detrimental to honeybees^{6–8} or obligate schooling fish^{1,2,13} and there is experimental evidence that such a trade-off also occurs in other animal groups where social learning of routes depends on the ability of informed individuals to maintain social interactions with those that are naive⁴.

Informed individuals within a group may differ in their preferred direction, however, due to differences in experience or motivation^{1,5–11}. Groups of animals often have to make collective decisions, such as to move together to a specific resource, such as a nest site^{1,5–9} or food source^{2,4}. The means by which such decisions can be made is very poorly understood, especially in the case of large groups and when individuals are not capable of knowing whether they are in a majority or minority, or how the quality of their information compares with that of others, or even whether there are any other individuals in the population with information. To investigate this we created two subsets of informed individual, each having its own directional preference.

If the number of individuals that exhibit each preference is equal, the direction of group motion depends on the degree to which the preferred preferences differ: as this difference increases, groups change from moving in the average preferred direction of all informed individuals to selecting randomly one of the two preferred

directions (Fig. 3A, a). If the number of individuals exhibiting one of the preferences is increased, however (Fig. 3A, b and c), the group will (given the appropriate difference in preference, Fig. 3A, a) select collectively the direction preferred by the majority, even if that majority is small (Fig. 3A, b).

This ability of a group to average preferences when differences are relatively small²¹ but to achieve a consensus for the majority option when differences are large is likely to be important in many animal groups, especially if the point at which this transition occurs can be tuned. To achieve this tuning we introduced a simple feedback on the weighting of the preferred direction ω . If, in a given time step, informed individuals find themselves moving in a similar direction (here within a 20-degree arc) to their preferred direction, ω is reinforced (by ω_{inc} up to a maximum, ω_{max}), otherwise it is reduced (by ω_{dec} to a minimum of 0). Such a feedback loop allows consensus decisions (as opposed to averaging) for smaller differences in preferred direction (Fig. 3B, C). Figure 3B, C shows that the transition to consensus decision-making occurs for small changes in the number of individuals in each subset and that the influence of the minority subset decreases rapidly as the difference in size of the informed subsets increases.

This decision-making mechanism also allows discrimination with respect to quality of information. For example, if there is no difference in the size of the two subsets of informed individual, but there exists a difference in their ability to correctly determine their

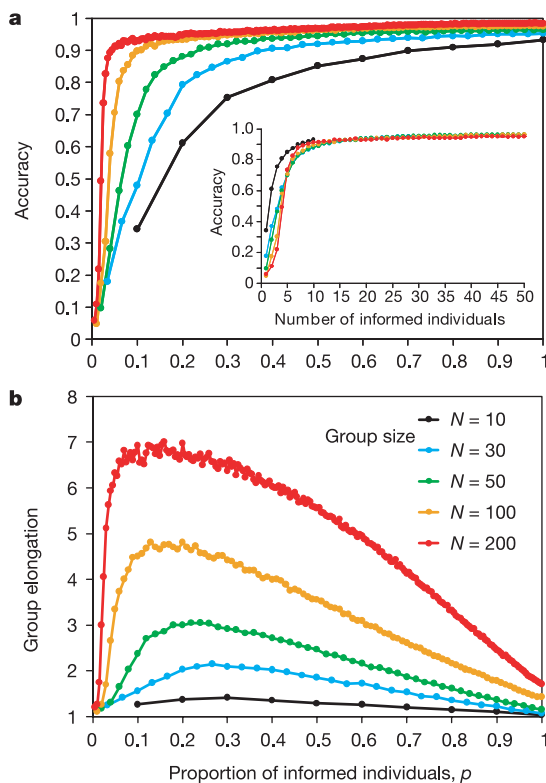


Figure 1 Group accuracy, and shape, as a function of the proportion of informed individuals p , for different group sizes N . The initial increase in accuracy as a function of p (a) is associated with the group becoming elongated as p increases (b). Informed individuals tend to occupy a frontal position within the group. The elongation of the group then decreases as p increases further and an increasingly large proportion of the group have knowledge of the directional vector $\mathbf{g}$. 400 replicates, $\omega = 0.5$; $\alpha = 1$, $\rho = 6$, $\gamma = 0$, $\Delta t = 0.2$ s, $\theta = 2$, $s_j = \alpha s^{-1}$ corresponding to fish^{2,19}.

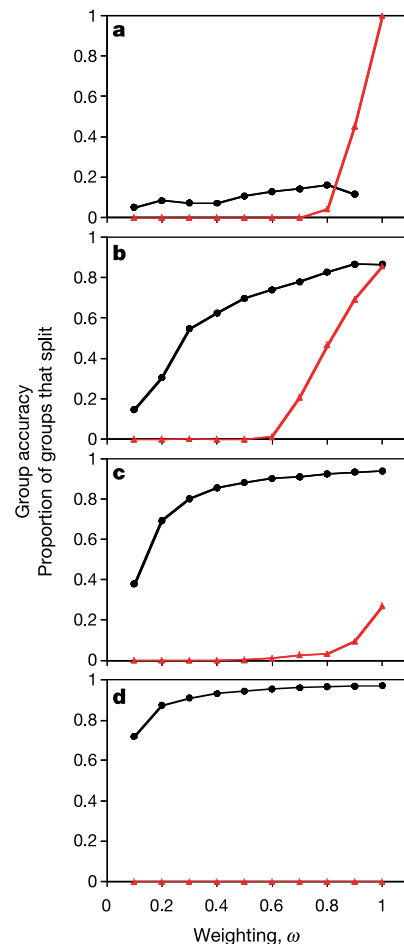


Figure 2 The influence of the weighting of preferred direction. The accuracy of group motion (black circles) and probability of group fragmentation (red triangles) as a function of weighting ω and the proportion of informed individuals p . a, $p = 0.02$ (1 individual); b, $p = 0.1$ (5 individuals); c, $p = 0.2$ (10 individuals); d, $p = 0.5$ (25 individuals). Parameters as for Fig. 1, $N = 50$.

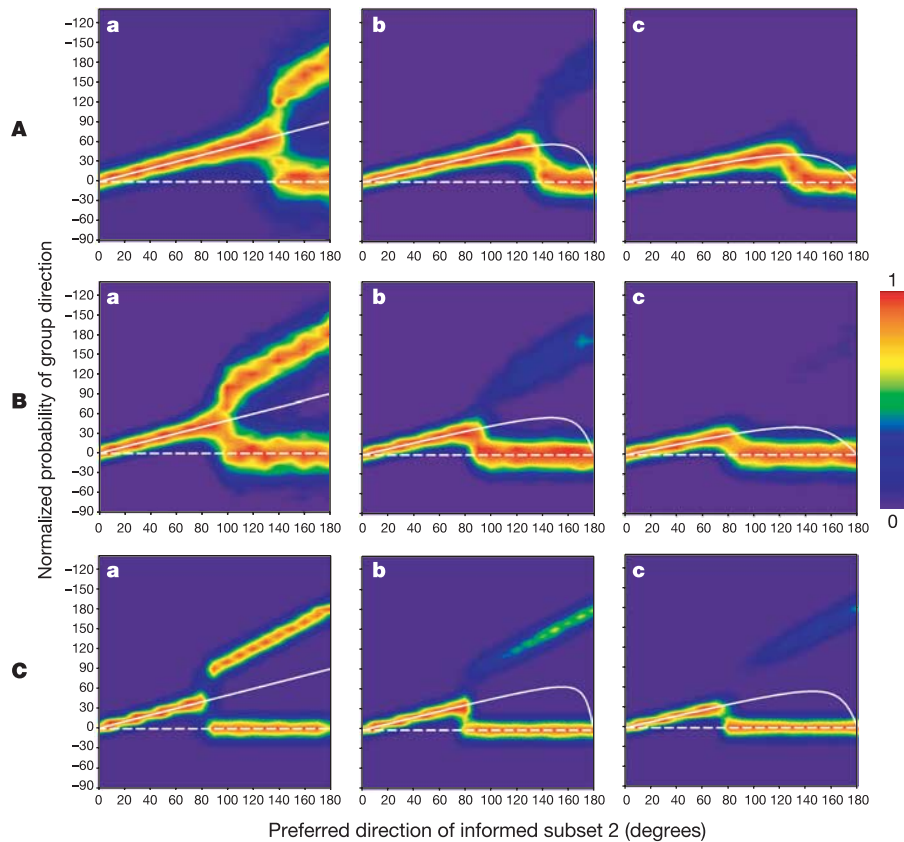


Figure 3 Collective selection of group direction when informed individuals differ in preference. Normalized probability distribution (proportion of maximum) of group direction for groups containing two subsets (with directional preferences s_1 and s_2 , and of sizes n_1 and n_2 , respectively) of informed individual within each group, each with its own directional preference: $s_1 = 0$ degrees (white dotted line) whereas $s_2 = 0-180$ degrees (2,000 replicates per 10-degree interval). Total group size $N = 100$ (parameters as for Fig. 1; $\omega, \omega_{\max} = 0.4$). We show the large influence of slightly changing n_1 and n_2 on group direction. In **a**, the first column, $n_1 = n_2$ (**A**, **B** = 5; **C** = 10, demonstrating

consistency at larger values of p). In **b**, the second column, n_1 is increased by 1, whereas in **c**, column 3, n_1 is increased by 1 and n_2 decreased by 1 (to allow direct comparison with the first column). **B** and **C** include feedback. **A**, **a**, $n_1 = 5, n_2 = 5$; **A**, **b**, $n_1 = 6, n_2 = 5$; **A**, **c**, $n_1 = 6, n_2 = 4$; **B**, **a-c**, n_1 and n_2 as for **A**, **a-c**, respectively. **C**, **a**, $n_1 = 10, n_2 = 10$; **C**, **b**, $n_1 = 11, n_2 = 10$. **C**, **c**, $n_1 = 11, n_2 = 9$. For **B** and **C** $\omega_{\text{inc}} = 0.012, \omega_{\text{dec}} = 0.0008$. Solid white lines are for reference only, representing the direction of the average vector $\mathbf{g}$ of all informed individuals (with constant ω).

preferred direction, the group can select collectively the direction associated with least error (Fig. 4, see Methods for details).

Our model demonstrates that efficient transfer of information, and decision-making, can occur within animal groups in the absence of explicit signals or complex mechanisms for information transfer. This means that informed and naive individuals do not have to be able to recognize each other and that leadership can emerge as a function of information differences among members of

a population, and is therefore transferable. No inherent differences between individuals (such as dominance due to larger body size) need to be invoked to explain leadership, although these properties can also influence group motion^{2,18}. Furthermore, the mechanism of coordination we propose here requires only limited cognitive ability, and demonstrates that individuals can respond spontaneously to those that have information. This is important to our understanding of group foraging, social learning, migration and navigation, and may provide new design protocols for information transfer among grouping robots. □

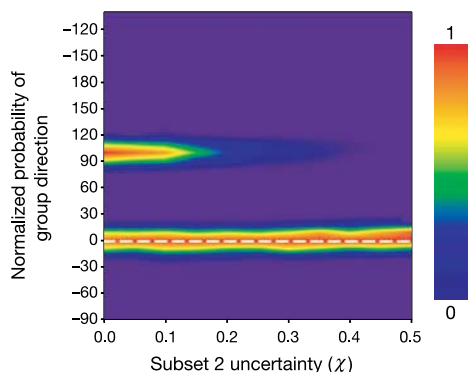


Figure 4 Discrimination between two directions ($s_1 = 0; s_2 = 100$ degrees) based on information quality (χ , see Methods). Parameters as for Fig. 3. $n_1 = 10, n_2 = 10$. Subset 2 has uncertain information when $\chi > 0$.

Methods

Motion in animals is subject to random influences (for example, sensory/movement error). This is simulated by modifying $\mathbf{d}_i(t + \Delta t)$ or $\mathbf{d}_i'(t + \Delta t)$, for uninformed and informed individuals, respectively, by rotating it by a random angle taken from a circular-wrapped gaussian distribution, centred on 0, with standard deviation $\sigma = 0.01$ radians (see ‘Uncertainty of information’, below) resulting in vector $\mathbf{d}_i''(t + \Delta t)$. Individuals can turn through an angle of, at most, $\theta \Delta t$ radians towards their desired direction in time Δt ; if the angle between $\mathbf{v}_i(t)$ and $\mathbf{d}_i''(t + \Delta t)$ is less than $\theta \Delta t$, they achieve alignment with their desired vector, $\mathbf{v}_i(t + \Delta t) = \mathbf{d}_i''(t + \Delta t)$, otherwise they turn $\theta \Delta t$ towards it. The new position vector of individual i is then given by $\mathbf{c}_i(t + \Delta t) = \mathbf{c}_i(t) + \mathbf{v}_i(t + \Delta t) \Delta t s_i$, where s_i is the speed of individual i .

Group size

Group sizes here are comparable to the size of schools, flocks or herds, of many species^{1-4,13,19}, but smaller than large aggregates such as honeybee colonies^{7,8}, owing to the nonlinear increase in computer processing time required as N increases. Our results, however, are likely to be independent of absolute group size, within the constraints of maintaining cohesion of group members²². To automatically test whether groups remained cohesive we used the equivalence class technique described in refs 18, 19.

Group direction

To quantify group direction $\mathbf{h}$ we create a vector extending from the group's centroid calculated at time $t_f \Delta t - 50 \Delta t$ to the centroid calculated at $t_f \Delta t$, where t_f the final time step, is 2,500. In Figs 1 and 2 we calculated the mean angular deviation s for 400 replicates, equivalent to calculating the linear standard deviation²⁰, which we normalized so that its minimum value is 0, corresponding to no information transfer (groups move in random directions), and its maximum value is 1, corresponding to the motion of the simulated groups always being exactly aligned with $\mathbf{g}$.

Elongation

Elongation was measured by creating a bounding box around the group aligned with the direction of travel and calculating the ratio of the length of the axis aligned with the group direction, to that perpendicular to group direction. This value is 1 when both axes are identical, >1 as the group becomes more elongated in the direction of travel, and <1 as it becomes elongated perpendicular to the direction of travel.

Uncertainty of information

Individuals may also not have perfect knowledge about their preferred direction $\mathbf{g}$, and this can be simulated for each individual i at the start of the simulation by rotating by the same type of circular-wrapped gaussian distribution with standard deviation γ , resulting in vector $\mathbf{g}_i$. Changing γ changes our results quantitatively, but not qualitatively, within the upper limits imposed by groups being able to maintain cohesion (see Supplementary Fig. 1).

To simulate a difference in the ability of informed individuals, within the same group, to correctly determine their preferred direction, $\mathbf{g}$ is rotated for $p/2$ individuals ($s/2$) by gaussian-distributed angle, with standard deviation χ radians (creating $\mathbf{g}_i$, as above) at the start of the simulation (see Fig. 4).

Starting conditions

Each simulation run was started with randomized individual positions and orientations.

Received 10 August; accepted 30 November 2004; doi:10.1038/nature03236.

- Krause, J. & Ruxton, G. D. *Living in Groups* 84–85, 137–143 (Oxford Univ. Press, Oxford, 2002).
- Couzin, I. D. & Krause, J. Self-organization and collective behaviour in vertebrates. *Adv. Study Behav.* **32**, 1–75 (2003).
- Reebs, S. G. Can a minority of informed leaders determine the foraging movements of a fish shoal? *Anim. Behav.* **59**, 403–409 (2000).
- Swaney, W., Kendal, J., Capon, H., Brown, C. & Laland, K. N. Familiarity facilitates social learning of foraging behaviour in the guppy. *Anim. Behav.* **62**, 591–598 (2001).
- Franks, N. R., Pratt, S. C., Mallon, E. B., Britton, N. F. & Sumpter, D. J. T. Information flow, opinion polling and collective intelligence in house-hunting social insects. *Phil. Trans. R. Soc. Lond. B* **357**, 1567–1583 (2002).
- Lindauer, M. Communication in swarm-bees searching for a new home. *Nature* **179**, 63–66 (1957).
- Seeley, T. D. *Honeybee Ecology: a Study of Adaptation in Social Life* 71–74 (Princeton Univ. Press, Princeton, 1985).
- Seeley, T. D. *The Wisdom of the Hive* 34–35 (Harvard Univ. Press, Cambridge, 1995).
- Seeley, T. D. Consensus building during nest-site selection in honey bee swarms: the expiration of dissent. *Behav. Ecol. Sociobiol.* **53**, 417–424 (2003).
- Conradt, L. & Roper, T. J. Group decision-making in animals. *Nature* **421**, 155–158 (2003).
- von Frisch, K. *The Dance Language and Orientation of Bees* 28–235 (Harvard Univ. Press, Harvard, 1967).
- Rubenstein, D. I. & Hack, M. Horse signals: the sounds of scents and fury. *Evol. Ecol.* **6**, 254–260 (1992).
- Partridge, B. L. Structure and function of fish schools. *Sci. Am.* **245**, 114–123 (1982).
- Berthold, P. & Querner, U. Genetic basis of migratory behaviour in European warblers. *Science* **212**, 77–79 (1981).
- Berthold, P., Helbig, A. J., Mohr, G. & Querner, U. Rapid microevolution of migratory behaviour in a wild bird species. *Nature* **360**, 668–670 (1992).
- Neill, W. H. Mechanisms of fish distribution in heterothermal environments. *Am. Zool.* **19**, 305–317 (1979).
- Grunbaum, D. Schooling as a strategy for taxis in a noisy environment. *Evol. Ecol.* **12**, 503–522 (1998).
- Couzin, I. D., Krause, J., James, R., Ruxton, G. D. & Franks, N. R. Collective memory and spatial sorting in animal groups. *J. Theor. Biol.* **218**, 1–11 (2002).
- Hoare, D. J., Couzin, I. D., Godin, J.-G. & Krause, J. Context-dependent group-size choice in fish. *Anim. Behav.* **67**, 155–164 (2004).
- Batschelet, E. *Circular Statistics in Biology* 34–36 (Academic, London, 1981).
- Simons, A. M. Many wrongs: the advantage of group navigation. *Trends Ecol. Evol.* **19**, 453–455 (2004).
- Gregoire, G., Chaté, H. & Tu, Y. Moving and staying together without a leader. *Physica D* **181**, 157–170 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements I.D.C. thanks the Pew Charitable Trusts, the NSF and the EPSRC for their support. I.D.C. and J.K. acknowledge an EPSRC grant and are also grateful for fellowships at the Centre for Interdisciplinary Research, University of Bielefeld, where we had the opportunity to develop this research. S.A.L. acknowledges support from the NSF and the Andrew W. Mellon Foundation, and N.R.F. from the EPSRC and the BBSRC. I.D.C. thanks Balliol College for support and S. Pratt, D. Rubenstein, D. James and A. Ward for their input.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.D.C. (iain.couzin@zoo.ox.ac.uk or icouzin@princeton.edu).

Insects breathe discontinuously to avoid oxygen toxicity

Stefan K. Hetz¹ & Timothy J. Bradley²

¹Department of Animal Physiology, Humboldt-Universität zu Berlin, Philippstr. 13, 10115 Berlin, Germany

²Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525, USA

The respiratory organs of terrestrial insects consist of tracheal tubes with external spiracular valves that control gas exchange. Despite their relatively high metabolic rate, many insects have highly discontinuous patterns of gas exchange, including long periods when the spiracles are fully closed. Two explanations have previously been put forward to explain this behaviour: first, that this pattern serves to reduce respiratory water loss¹, and second, that the pattern may have initially evolved in underground insects as a way of dealing with hypoxic or hypercapnic conditions². Here we propose a third possible explanation based on the idea that oxygen is necessary for oxidative metabolism but also acts as a toxic chemical that can cause oxidative damage of tissues even at relatively low concentrations. At physiologically normal partial pressures of CO₂, the rate of CO₂ diffusion out of the insect respiratory system is slower than the rate of O₂ entry; this leads to a build-up of intratracheal CO₂. The spiracles must therefore be opened at intervals to rid the insect of accumulated CO₂, a process that exposes the tissues to dangerously high levels of O₂. We suggest that the cyclical pattern of open and closed spiracles observed in resting insects is a necessary consequence of the need to rid the respiratory system of accumulated CO₂, followed by the need to reduce oxygen toxicity.

The respiratory system of insects consists of a highly branched system of cuticle-lined tubes extending throughout the body³. The tubules are filled with air, which greatly facilitates the transport of gases through the body (the diffusion of O₂ and CO₂ is about 10⁶ and 10⁴ times faster, respectively, in air than in water, blood or tissue)⁴. The tracheae open to the external atmosphere through valve-like spiracles on the surface of the abdomen and thorax. Internal to the spiracles are large tracheal trunks and distensible air sacs. Smaller tracheae extend from the tracheal trunks in a branching, dendritic pattern. The finest branches at the tips of the tracheal system, termed tracheoles, can be less than a micrometre in diameter. They lie adjacent to the tissues and serve as the principal site for gas exchange.

Early calculations suggested that passive diffusion of gases in the tracheal system of insects should be more than adequate to support oxidative metabolism, even in fairly large insects⁴. Observation of living insects reveals, however, that they have elaborate processes for changing the hydrostatic pressure in the haemocoel, thereby actively ventilating the tracheal system^{5–7}. Even more surprisingly, many insects show highly rhythmic patterns of spiracular control (the discontinuous gas exchange cycle, DGC) where the tracheal system is periodically completely closed, followed by lengthy periods during which respiratory exchange is severely constricted^{8–10}. It is intriguing to insect physiologists that a system which can apparently be operated passively instead exhibits a complex system of controls that are both metabolically costly and would seem to interfere with, rather than foster, respiratory homeostasis.

The discontinuous gas exchange cycle of insects is characterized by a period in which the spiracles are fully closed (the 'closed phase', Fig. 1). During this time, the partial pressure of oxygen (p_{O_2}) in the lumina of the tracheae declines, and the partial pressure of CO₂ (p_{CO_2}) increases. Owing to the high solubility of CO₂ in the aqueous haemolymph and a respiratory quotient (the amount of CO₂

produced divided by the O_2 consumed) that is normally less than one in diapausing pupae, the increase in p_{CO_2} in the air within the tracheae is less than the decrease in p_{O_2} . As a result, the total atmospheric pressure in the tracheal lumina declines slightly during the closed phase. Eventually, the spiracles begin to flutter, allowing bulk flow of air down this pressure gradient. After a period of time, a third phase begins in which the spiracles remain fully open (the 'open phase'). During this period, both p_{O_2} and p_{CO_2} tend towards levels present in the external atmosphere due to the diffusion of these gases down their concentration gradients through the tracheae and open spiracles. It has been argued that the 'flutter phase' is initiated by critically low p_{O_2} in the tracheae, and the open phase is initiated by critically high p_{CO_2} (ref. 1).

Although the phenomenon of discontinuous gas exchange has been extensively studied in insects, its adaptive significance is a subject of considerable debate. Two models have been put forward. The first argues that the evolutionary force promoting discontinuous ventilation is the reduction of respiratory water loss occasioned by the closed phase and bulk inward flow of air during the flutter phase¹. Subsequent studies have failed to find much support for the idea that the fitness benefits of the pattern lie in water conservation^{2,10,11}. Although many insects show discontinuous ventilation, most do not. Insects show discontinuous ventilation only when they are not active or moving. Insects such as grasshoppers that can be exposed to high temperatures and dry air during the day do not show discontinuous ventilation at such times, but do show discontinuous ventilation at night when respiratory water loss is already minimized¹⁰. Fruitflies that have undergone selection for enhanced desiccation resistance show genetically determined modifications in their respiratory patterns¹². Those flies that showed discontinuous ventilation did not have reduced rates of water loss compared with control flies that did not breathe discontinuously. Similarly, individual flies had identical rates of water loss during discontinuous breathing periods as in periods of regular breathing¹².

A second explanation for the occurrence of discontinuous gas exchange in insects has recently been proposed². Many of the insects showing discontinuous gas exchange spend portions of their life cycles underground. The DGC can be beneficial in an environment in which p_{O_2} is low and p_{CO_2} is high. For example, partial closing of the spiracles promotes a low p_{O_2} in the tracheal lumina, allowing inward diffusion in hypoxic environments. Partial closing of the spiracles also promotes the accumulation of CO_2 in the tissues and tracheae, promoting the rapid release of CO_2 during the open phase

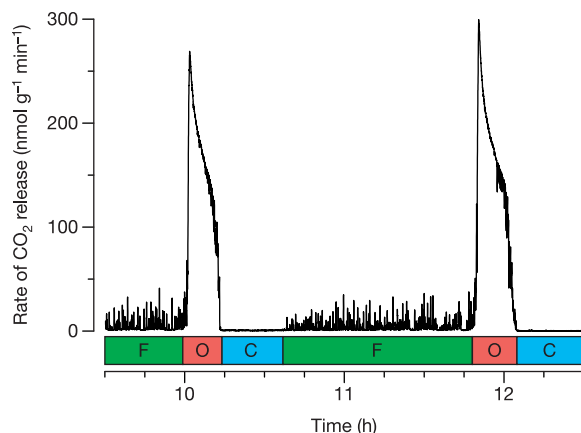


Figure 1 The rate of release of CO_2 from a pupa of *Attacus atlas* over time. A burst of CO_2 release is observed during the open phase (O, red bar). During the closed phase (C, blue bar), the spiracles are closed and CO_2 release is low. The closed phase is followed by a flutter phase (F, green bar) during which CO_2 release occurs only during brief intervals of spiracular opening.

in hypercapnic environments. Finally, the discontinuous nature of the respiratory exchange allows for the diffusion of gases surrounding the animal in an environment where convective gas exchange is limited. A number of the insects showing discontinuous gas exchange do spend substantial portions of their life cycles underground, including ants, lepidopteran pupae (such as those of sphingid moths) and fossorial beetles^{11,13}. For other insects exhibiting discontinuous gas exchange, including grasshoppers and triatomid hemipterans^{10,14}, the connection to underground living is more tenuous. Although the DGC unquestionably provides advantages in a hypoxic or hypercapnic environment, it is harder to imagine the evolutionary pathway by which this respiratory pattern arose. Our understanding of the evolution of the insect orders does not suggest that the ancestor of all of the insects exhibiting discontinuous gas exchange was fossorial. Instead, it may be that features of the insects' respiratory control systems promote discontinuous ventilation when faced with the conditions which prevail underground.

We propose that an additional control process is responsible for discontinuous ventilation in insects. We suggest that O_2 regulation

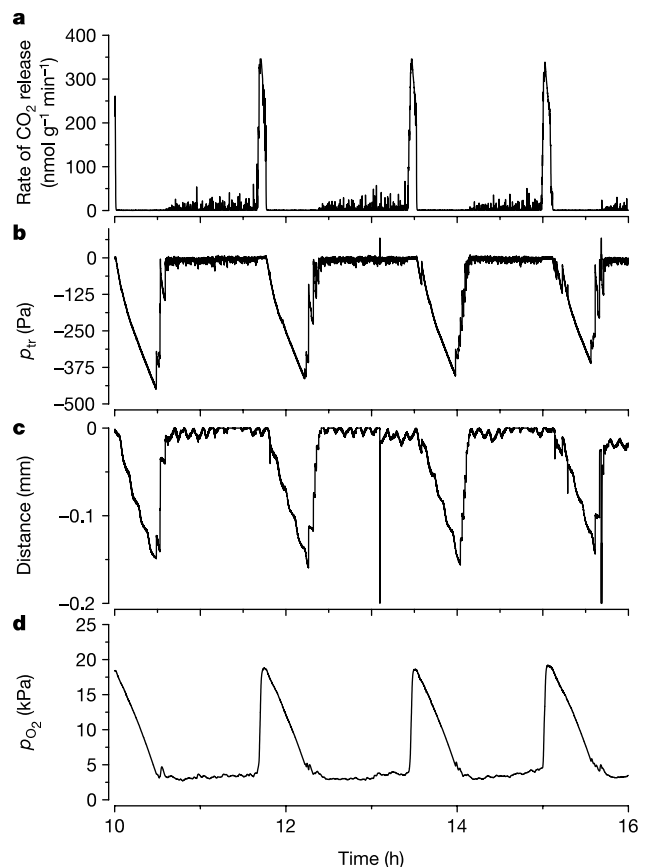


Figure 2 The discontinuous gas exchange cycle (DGC) in a pupa of *A. atlas*. **a**, The discontinuous release of CO_2 from the insect into the air stream flowing through the respirometer. The large spikes represent an open phase when the spiracles are open. Subsequent periods of low CO_2 releases reflect the closed phase. Periods with small spikes of CO_2 release reflect the flutter phase. **b**, Air pressure in the tracheae relative to atmospheric pressure. The intratracheal pressure (p_{tr}) is reduced below atmospheric pressure only during the closed phase. **c**, The distance between a marker placed on the tip of the abdomen and a sensor on the chamber wall sensing distance to the marker. A shortening of the abdomen is indicated as a downward trend in the values. The abdomen shortens during each closed phase owing to the increase in negative pressure within the tracheae. **d**, Intratracheal p_{O_2} . During the open phase, p_{O_2} in the tracheae approaches that in the atmosphere (20.4 kPa). During the closed phase, p_{O_2} drops to a value around 4–5 kPa. The flutter phase serves to keep p_{O_2} near this level.

(in fact, the intentional exclusion of O_2) drives the control of spiracles and ventilatory physiology in insects. It is this O_2 regulation that produces the observed discontinuous release of CO_2 .

During the open phase of the DGC, O_2 rapidly enters the tracheae, diffusing into the cells and bringing them to a higher p_{O_2} (Fig. 2). The respiratory system then enters the closed phase. We suggest that the purpose of this closed phase is specifically to reduce p_{O_2} around the cells to a lower, and safer, level. As p_{O_2} declines and eventually reaches a critical lower concentration, fluttering begins to regulate p_{O_2} . The diffusive resistance of the tracheal system has been determined for the goat moth *Cossus*⁴. When the spiracles are fully open, the p_{O_2} at the tips of the tracheoles should be about 19 kPa. The values shown in Fig. 2 verify this estimate. This is a very high level of oxygen for tissues to be exposed to. For example, in vertebrates, the p_{O_2} in the capillaries of inactive tissues is about 5 kPa; active tissues such as contracting muscle experience p_{O_2} levels of 0.5 kPa in the capillaries and 0.4 kPa in the cells themselves¹⁵. Oxygen radicals are produced in mitochondria and microsomes in an O_2 -dependent manner¹⁶ and normoxia has been shown to induce extensive oxidative damage both *in vivo* and *in vitro*¹⁷.

When the spiracles of insects are open, therefore, the tissues are exposed to a p_{O_2} that is unusual among biological systems and that can lead to dangerous levels of oxidative damage. In the fruitfly *Drosophila melanogaster*, oxidative damage accumulates in the tissues with time even under normal atmospheric conditions^{18,19}.

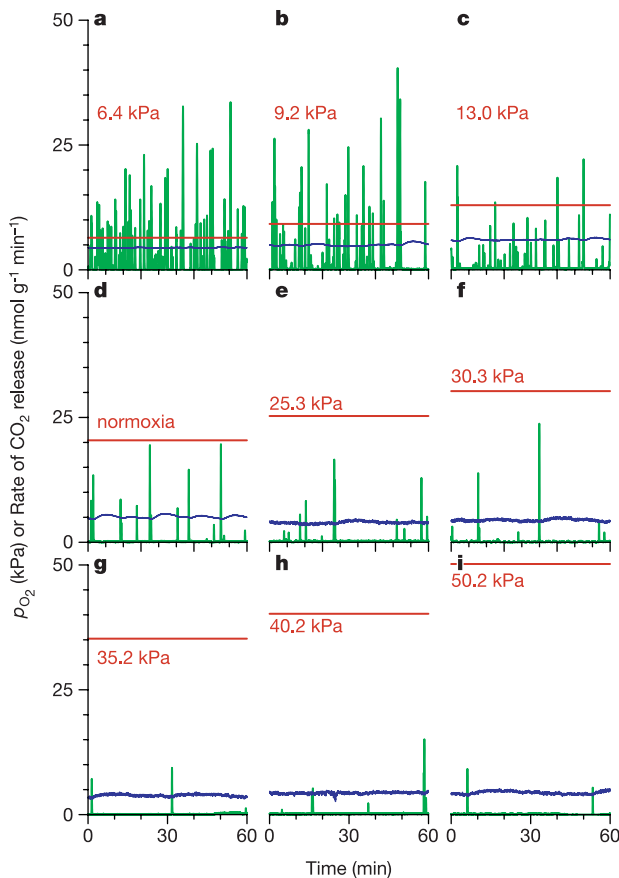


Figure 3 Respiratory regulation in pupae of *A. atlas* during the flutter phase at different atmospheric O_2 concentrations. Green lines indicate the rate of CO_2 release by the insect. Red lines and red numbers indicate the ambient p_{O_2} . Blue lines indicate the measured p_{O_2} in the tracheae of the insects. **a–c**, Respiratory patterns under conditions of hypoxia. More CO_2 is released during the flutter phase under hypoxic conditions because the insects open the spiracles more frequently to obtain sufficient O_2 . **d**, CO_2 release under conditions of normoxia. **e–i**, Under conditions of hyperoxia, the flutter period is almost eliminated. Intratracheal p_{O_2} is maintained at about 4 kPa regardless of the external p_{O_2} .

In the literature on ageing, this is seen as a deleterious effect of normal O_2 concentrations on genetically normal flies, a process which contributes to premature death. Flies engineered to over-express enzymes that reduce oxidative damage, such as catalase, Cu-Zn superoxide dismutase or Mn superoxide dismutase, exhibit longer life and postponed ageing^{18,20,21}. This oxidative damage is influenced by the level of O_2 present, shown by the fact that increases in the level of O_2 in the atmosphere that the insects are breathing lead to enhanced rates of oxidative damage and reduced longevity²². Oxygen has been shown to be a toxic molecule that is needed for oxidative metabolism but must be supplied in carefully controlled amounts and concentrations¹⁷.

By using a closed phase and then transitioning into the flutter phase, insects first rapidly reduce O_2 levels near the tissues (Fig. 2) and then provide O_2 at a rate that supports oxidative metabolism (Fig. 3). Note that the 'water-saving' and 'underground' models for respiratory control would both predict that O_2 concentrations should be excessively high under conditions of hyperoxia. This is not the case (Fig. 3). Instead the insects show oxygen regulation at all external O_2 concentrations.

In an 'ideally' designed respiratory system, insects would obtain O_2 and lose CO_2 at similar rates; that is, with the same stoichiometry as seen for O_2 uptake and CO_2 production during aerobic metabolism (respiratory quotient varies from 0.7 to 1.0). This is impossible in insects because O_2 and CO_2 move in opposite directions along an identical pathway. The partial pressures of O_2 are 20.9 kPa in the external atmosphere and 6 kPa in the lower reaches of the tracheae, a gradient of 15 kPa. The peak partial pressures of CO_2 at the end of the flutter phase are around 6.5 kPa, whereas the outside pressure is effectively zero, a gradient of 6.5 kPa in the opposite direction. As a consequence of the gradients (and the very slight differences in rate of diffusion due to differences in molecular weight), O_2 will always enter the insects more rapidly than CO_2 leaves.

The substantial restriction of the spiracles, with occasional opening to permit O_2 entry, serves to regulate p_{O_2} at the tissue level over a wide range of ambient partial pressures of O_2 (Fig. 4). This regulation by necessity leads to the accumulation of CO_2 in the tissues, haemolymph and tracheae. When this accumulation reaches a critical level, the insects have no choice but to open their spiracles and allow the CO_2 to escape. As a result, the insects receive O_2 more rapidly than they need or want when the spiracles are open. When the need for CO_2 release has been met, the system closes completely to reduce the high levels of O_2 . Cycling is therefore an inevitable aspect of the insect respiratory system given its physical design and control mechanisms. The fact that some insects do not exhibit discontinuous gas exchange, in either a time-dependent or species-specific manner, can be attributed to the presence of a high

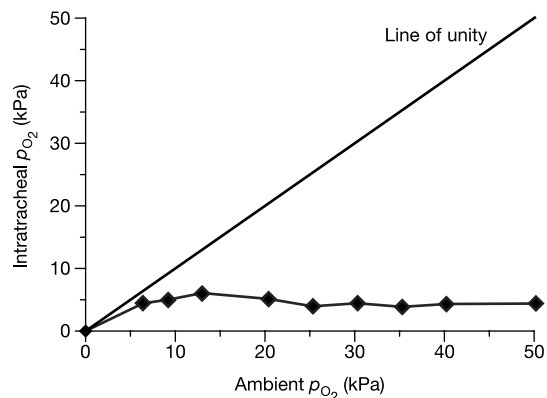


Figure 4 Intratracheal p_{O_2} in pupae of *A. atlas* during the flutter phase as a function of external O_2 concentration. The respiratory system maintains intratracheal p_{O_2} at a low level regardless of external, ambient p_{O_2} .

metabolic rate relative to the maximum O₂ delivery capacity of the respiratory system. Discontinuous ventilation is observed in some non-insect, tracheate arthropods (such as sun scorpions, solifugids and ticks)^{23–25}. We propose that these arthropods must also reduce gas exchange during periods of low metabolic activity to avoid the toxic effects of oxygen.

One can quite reasonably ask, given the capacity of evolutionary forces to refine and improve physiological systems, why the insect respiratory system is designed in a manner that subjects the insects to periodic bursts of toxic O₂ levels and complicates respiratory and pH homeostasis. The answer lies in the fact that the DGC is observed only in resting insects²⁶. The insect's respiratory system has been designed to function most efficiently at high levels of O₂ consumption. We argue that the insects' respiratory system is designed to provide a high delivery capacity, not to swamp the tissue with oxygen during periods of low aerobic demand but rather to supply abundant oxygen during periods of high oxygen demand. Extensive studies of the design of the vertebrate respiratory system have shown that its design is driven by demands during periods of maximum aerobic activity^{27,28}. The respiratory system of insects is similarly 'overdesigned' for providing oxygen to the tissues during non-exercise conditions. The respiratory pattern that we observe during low demand is the insect's attempt to use a high capacity system during periods of 'metabolic idling'.

Evidence in support of this notion is provided by the fact that the cycle disappears when insects increase their metabolic rate, for example in high temperatures or during exercise^{26,29}. Under these conditions the cells use O₂ at a much faster rate. We observe that the spiracles do not fully close under these conditions by the fact that the release of CO₂ never goes to zero. We propose that an O₂-sensing system is located somewhere in the animal and that under conditions of enhanced metabolic rate, this permits the spiracles to remain more fully open. Previously, physiologists have reasoned that the spiracles open during exercise to provide more oxygen. We would argue that they open because they are released from their need to close. □

Methods

To measure the hydrostatic pressure and the O₂ pressure within the tracheal system, two spiracles were intubated. A short length of polyethylene tubing (length 50 mm, outer diameter 0.9 mm; inner diameter 0.4 mm) was inserted past the spiracular valve into one spiracle in order to record the intratracheal pressure. The spiracle was then carefully sealed with wax. The tubing was fitted to a miniature differential pressure sensor (SenSym SDXL010D4, Sensortronics) that was connected to a custom-made amplifier. The pressure range was 2,500 Pa, with an accuracy of about ±10 Pa. Abdominal movements were recorded optically with a miniature infrared light reflection switch (SFH900, Siemens). The light reflected from a reflecting sheet (23090, 3M) glued to the tip of the abdomen was used to record the distance between the sensor and the tip of the abdomen with a custom-made amplifier built to obtain a linear relation of output voltage versus distance. Tracheal pO₂ was measured with a custom-made Clark-type pO₂ electrode through a single spiracle as previously described³⁰. The O₂ sensor was calibrated before and after experiments using moistened air and pure nitrogen at 15 °C. The animal (with sensors) was housed in a flow-through, temperature-controlled respirometer at 15 °C and CO₂ release patterns were recorded using a differential infrared gas analyser (URAS 4, Hartmann & Braun). Ambient pO₂ was modified using a computer-controlled mass-flow controller (MKS 1259, 200 ml min⁻¹) that permitted the addition of nitrogen or oxygen to the ambient air. Accuracy of the mixtures was checked by means of a flow-through O₂ sensor (Ametek S-3A/II with Sensor N-37).

Received 3 June; accepted 6 October 2004; doi:10.1038/nature03106.

- Levy, R. I. & Schneiderman, H. A. Discontinuous respiration in insects. II. The direct measurement and significance of changes in tracheal gas composition during the respiratory cycle of silkworm pupae. *J. Insect Physiol.* **12**, 83–104 (1966).
- Lighton, J. R. B. Notes from the underground: toward ultimate hypotheses of cyclic, discontinuous gas-exchange in tracheate arthropods. *Am. Zool.* **38**, 483–491 (1998).
- Wigglesworth, V. B. *The Principles of Insect Physiology* (Methuen, London, 1970).
- Krogh, A. Studien ueber tracheenrespiration. II. Ueber gasdiffusion in den tracheen. *Pfluegers Arch.* **179**, 95–112 (1920).
- Krogh, A. Studien ueber die tracheenrespiration. III. Die kombination von mechanischer ventilation mit gasdiffusion nach versuchen an dytiscidenlarven. *Pfluegers Arch.* **179**, 113–120 (1920).
- Slama, K. A new look at insect respiration. *Biol. Bull.* **175**, 289–300 (1988).
- Westneat, M. W. *et al.* Tracheal respiration in insects visualized with synchrotron x-ray imaging. *Science* **299**, 558–560 (2003).
- Slama, K. Active regulation of insect respiration. *Ann. Entomol. Soc. Am.* **92**, 916–929 (1999).

- Lighton, J. R. B. Discontinuous gas exchange in insects. *Annu. Rev. Entomol.* **41**, 309–324 (1996).
- Hadley, N. F. Ventilatory patterns and respiratory transpiration in adult terrestrial insects. *Physiol. Zool.* **67**, 175–189 (1994).
- Chown, S. L. & Holter, P. Discontinuous gas exchange cycles in *Aphodius fossor* (Scarabaeidae): a test of hypotheses concerning origins and mechanisms. *J. Exp. Biol.* **203**, 397–403 (2000).
- Williams, A. E. & Bradley, T. J. The effect of respiratory pattern on water loss in desiccation-resistant *Drosophila melanogaster*. *J. Exp. Biol.* **201**, 2953–2959 (1998).
- Lighton, J. R. B. & Garrigan, D. Ant breathing: testing regulation and mechanism hypotheses with hypoxia. *J. Exp. Biol.* **198**, 1613–1620 (1995).
- Bradley, T. J., Brethorst, L., Robinson, S. & Hetz, S. Changes in the rate of CO₂ release following feeding in the insect *Rhodnius prolixus*. *Physiol. Biochem. Zool.* **76**, 302–309 (2002).
- Richardson, R. S., Noyszeski, E. A., Kendrick, K. F., Leigh, J. S. & Wagner, P. D. Myoglobin O₂ desaturation during exercise: evidence of limited O₂ transport. *J. Clin. Invest.* **96**, 1916–1926 (1995).
- Jamieson, D. Oxygen toxicity and reactive oxygen metabolites in mammals. *Free Radic. Biol. Med.* **7**, 87–108 (1989).
- Fridovich, I. Oxygen is toxic! *Bioscience* **27**, 462–466 (1977).
- Sohal, R. S., Agarwal, S. & Orr, W. C. Simultaneous overexpression of copper- and zinc-containing superoxide dismutase and catalase retards age-related oxidative damage and increases metabolic potential in *Drosophila melanogaster*. *J. Biol. Chem.* **270**, 15671–15674 (1995).
- Schwarze, S. R., Weindruch, R. & Aiken, J. M. Oxidative stress and aging reduce Cox 1 RNA and cytochrome oxidase activity in *Drosophila*. *Free Radic. Biol. Med.* **25**, 740–747 (1998).
- Orr, W. C. & Sohal, R. S. Extension of life-span by over-expression of superoxide dismutase and catalase in *Drosophila melanogaster*. *Science* **263**, 1128–1130 (1994).
- Sun, J., Folk, D., Bradley, T. J. & Tower, J. Induced overexpression of mitochondrial Mn-superoxide dismutase extends the life span of adult *Drosophila melanogaster* without changing metabolic rate. *Genetics* **161**, 661–672 (2002).
- Yan, L. J. & Sohal, R. S. Mitochondrial adenine nucleotide translocase is modified oxidatively during aging. *Proc. Natl Acad. Sci. USA* **94**, 11168–11172 (1997).
- Lighton, J. R. B. & Fielden, L. J. Gas exchange in wind spiders (Arachnida, Solpugidae). Independent evolution of convergent control strategies in solpugids and insects. *J. Insect Physiol.* **42**, 347–357 (1996).
- Fielden, L. J. & Lighton, J. R. B. Effects of water stress and relative humidity on ventilation in the tick *Dermacentor andersoni* (Acari, Ixodidae). *Physiol. Zool.* **69**, 599–617 (1996).
- Lighton, J. R. B. & Joos, B. Discontinuous gas exchange in the pseudoscorpion *Garypus californicus* is regulated by hypoxia, not hypercapnia. *Physiol. Biochem. Zool.* **75**, 345–349 (2002).
- Lighton, J. R. B. Discontinuous CO₂ emission in a small insect, the formicine ant *Camponotus vicinus*. *J. Exp. Biol.* **134**, 363–376 (1988).
- Weibel, E. R. Symmorphosis and optimization of biological design: introduction and questions. In *Principles of Animal Design* (Cambridge Univ. Press, Cambridge, 1998).
- Weibel, E. R. *et al.* Design of the oxygen and substrate pathways. 7. Different structural limits for oxygen and substrate supply to muscle mitochondria. *J. Exp. Biol.* **199**, 1699–1709 (1996).
- Davis, A. L. V., Chown, S. L. & Scholtz, C. H. Discontinuous gas-exchange cycles in *Scarabaeus dung* beetles (Coleoptera: Scarabaeidae): Mass-scaling and temperature dependence. *Physiol. Biochem. Zool.* **72**, 555–565 (1999).
- Hetz, S. K., Wasserthal, L. T., Hermann, S., Kaden, H. & Oelssner, W. Direct oxygen measurements in the tracheal system of Lepidopterous pupae using miniaturized amperometric sensors. *Bioelectrochem. Bioenerg.* **33**, 165–170 (1994).

Acknowledgements We thank N. Heisler for comments and for providing equipment. T.J.B. would like to thank N. Heisler for his hospitality during a research visit. This work was supported by an NSF grant to T.J.B.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.J.B. (tbradley@uci.edu).

Asymptotic prey profitability drives star-nosed moles to the foraging speed limit

Kenneth C. Catania & Fiona E. Remple

Department of Biological Sciences, Vanderbilt University, VU Station B, Box 351634, Nashville, Tennessee 37235, USA

Foraging theory provides models for predicting predator diet choices assuming natural selection has favoured predators that maximize their rate of energy intake during foraging^{1–6}. Prey profitability (energy gained divided by prey handling time) is an essential variable for estimating the optimal diet. Time constraints of capturing and consuming prey generally result in

handling times ranging from minutes to seconds, yet profitability increases dramatically as handling time approaches zero, providing the potential for strong directional selection for increasing predator speed at high encounter rates (tiny increments in speed increase profitability markedly, allowing expanded diets of smaller prey). We provide evidence that the unusual anatomical and behavioural specializations characterizing star-nosed moles resulted from progressively stronger selection for speed, allowing the progressive addition of small prey to their diet. Here we report handling times as short as 120 ms (mean 227 ms) for moles identifying and eating prey. 'Double takes' during prey identification suggest that star-nosed moles have reached the speed limit for processing tactile information. The exceptional speed of star-nosed moles, coupled with unusual specializations for finding and eating tiny prey, provide new support for optimal foraging theory.

In the classical prey selection model for predicting predator diets³ the rate of energy intake (R) is equivalent to $E/(T_s + T_h)$, where E is the energy obtained from prey, T_s is the time spent searching for prey, and T_h is the time spent handling prey^{4–7}. Thus, minimizing T_s and T_h maximizes R . For most species there are numerous constraints on how short search and handling times can become. Yet the pay-offs for energy intake could rise dramatically if the denominator ($T_s + T_h$) is minimized, providing the potential for progressively stronger directional selection for increasing predator speed if short search and handling times can be achieved. The result would presumably be extreme sensory-motor specializations. Investigations of such species (that is, those species that have been pushed to the speed limits for foraging behaviours) may provide new insights into predator evolution. Star-nosed moles seem to have reached this extreme. Star-nosed moles have long been considered a biological enigma, exhibiting unusual anatomical and behavioural specializations that set them apart from other mammals⁸. Here we report that they are also the fastest of mammalian foragers, and suggest that their unusual anatomy and behaviour are the result of unusually strong directional selection for increasing predator speed,

allowing them to exploit small invertebrates in their wetland habitat⁹.

Using a high-speed video system we documented star-nosed moles detecting and eating small prey items as they foraged on a glass plate in a simulated tunnel. Moles searched using their mechanosensory star with a stereotyped behaviour that can be divided into a search phase and a feeding phase. During the search phase, the star was rapidly touched to different areas in the environment (Fig. 1). The average duration of the movements and touches was approximately 50 and 25 ms, respectively¹⁰. Thus a foraging mole typically contacted 13 different areas with the star every second. Contact of the star to prey constitutes the beginning of an 'encounter' and usually resulted in a foveation movement that brought the high-resolution 11th appendage pair—the tactile fovea⁸—into contact with the prey item, followed by capture of the prey with small, tweezer-like front teeth.

Figure 1a shows the relationship of the search phase to the feeding phase. Handling times were calculated to reflect the "principle of lost opportunity"⁶; that is, foraging time lost to consume prey. Thus, handling time began at the end of the first touch and lasted until the mole had eaten and began a star movement, resuming the search phase. Here we have considered handling times for the smallest prey that were detected and eaten (pieces of earthworm, 0.8–2.0 mm diameter, with an average energy value of approximately 10 J^{11}). The (weighted) average handling time for three moles was 227 ms for these prey sizes (mole 1: 228 ms, 85 trials; mole 2: 223 ms, 40 trials; mole 3: 230 ms, 24 trials). Figure 1 shows selected frames from high-speed video of moles encountering and eating prey (see also Supplementary Video 1). The handling time ranged from 120 to 440 ms.

Prey profitability (P)—the ratio of energy gained to handling time for each prey item (E/T_h)—provides a useful index of value for each item and is an important parameter for determining whether a prey type should be included in the optimal diet^{3–7}. The short handling times for star-nosed moles have a marked effect on profitability of small prey (Fig. 2a). Remarkably, tiny prey items

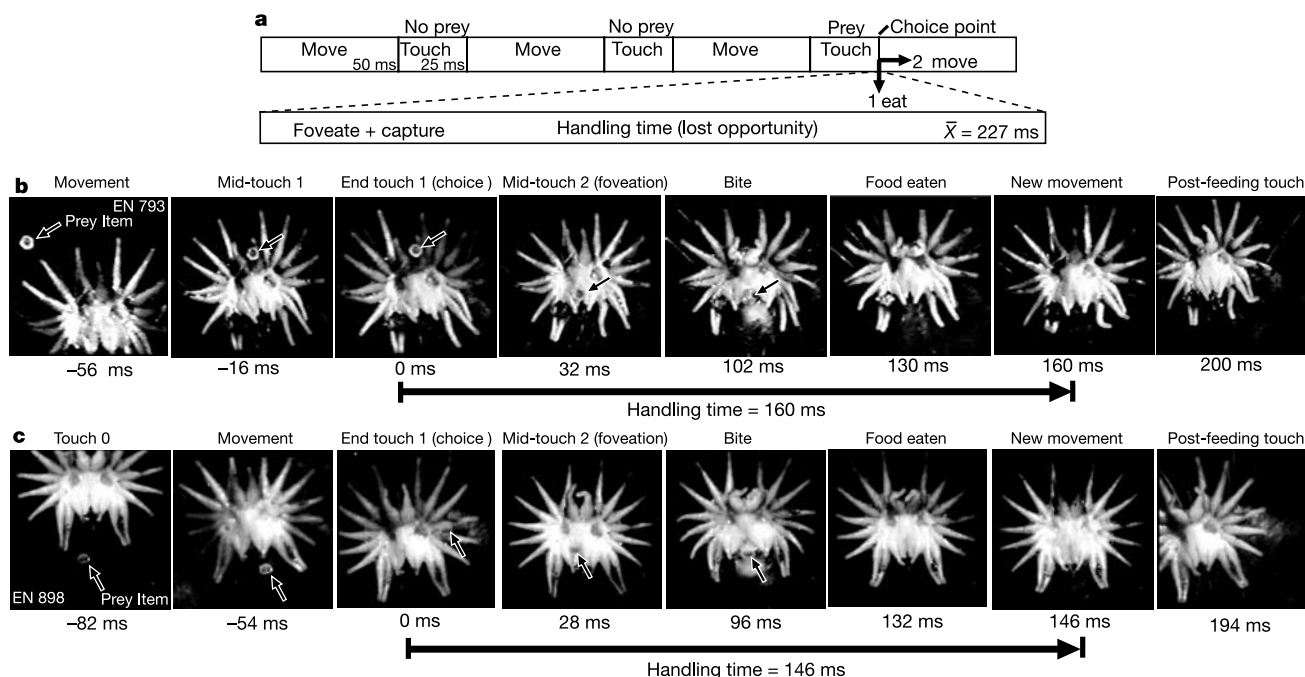


Figure 1 Star-nosed mole foraging. **a**, Search behaviour consisting of repeated cycles of star movements and touches lasting an average of 50 and 25 ms, respectively. When prey is encountered, moles feed (choice 1) or continue searching (choice 2). Feeding includes a foveation movement followed by grasping prey. **b**, **c**, Selected video frames illustrating

two feeding responses and the handling times. Handling time begins at the end of the first touch to prey and continues until searching is resumed. Arrows mark prey item (piece of earthworm).

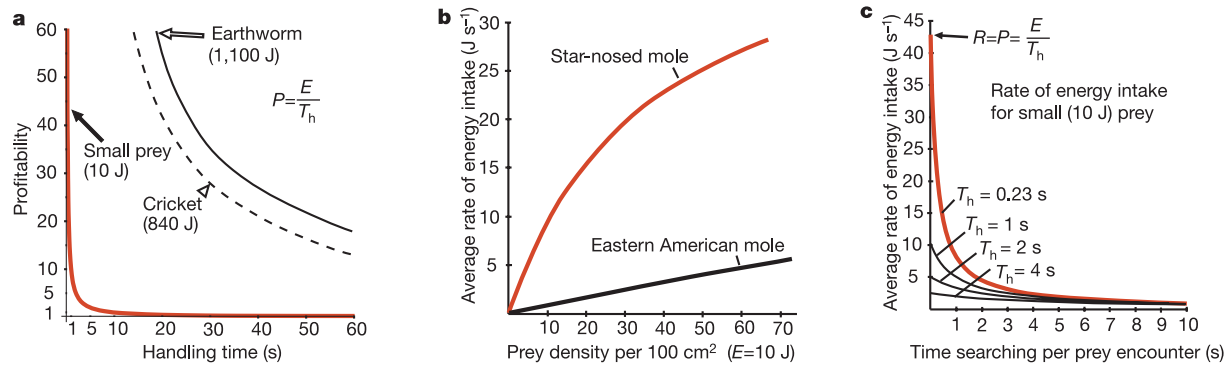


Figure 2 Prey profitability and energy intake for different handling times and prey densities. **a**, Red line represents small prey profitability. Short handling times (230 ms) make small prey profitable (filled arrow). Short-tailed shrews (dotted line) took 30 s to eat 840-J crickets (open arrowhead) that star-nosed moles avoid. Larger earthworm segments (1,100 J) were more profitable to star-nosed moles ($T_h = 19 s$), but their expected lower density suggests that moles would include small prey in their diet⁶. **b**,

Predicted effect of search speed on average energy intake for star-nosed moles and slower eastern moles with smaller receptor arrays, assuming equal T_h and small prey (10 J) (see text). **c**, The effect of T_h on average energy intake for different lengths of search between small prey encounters. Red line indicates star-nosed moles. Note that the maximum possible energy intake (arrow) is equivalent to prey profitability³ ($43 J s^{-1}$ for star-nosed moles).

may be more profitable to star-nosed moles than large insects (open arrowhead, Fig. 2a) are to other, competing species (Supplementary Video 2). Even large earthworm segments have only a slightly greater profitability to star-nosed moles than the tiny prey items (Fig. 2a; see also Supplementary Video 3). These considerations suggest that small prey should be included in the star-nosed mole's diet, as large prey would have to be common (encountered once every 10 s in this simplified example, see ref. 6 for calculating diet breadth) for small prey to be rejected. In fact, the star-nosed mole's unusual dentition provides evidence for a long history of selective pressure for ingesting tiny prey (Fig. 3a, b). The unique incisors function like a pair of tweezers that are coordinated with the final movements of the star to rapidly grasp tiny prey (Fig. 1b, c frame 5; see also Supplementary Videos). This observation is consistent with gut content analysis of star-nosed moles, which indicates that they consume large numbers of small invertebrates⁹.

Clearly star-nosed moles are specialized to minimize handling time (T_h) for small prey. However, to maximize energy intake the encounter rate with prey must also be maximized (thus minimizing T_s). Habitat choice has an obvious role in the type and number of prey that are encountered, and the star-nosed mole's unique (for moles) wetland habitat contains high densities of invertebrates^{9,12}. In addition, encounter rates can be increased by searching quickly and increasing the size of the sensory receptor sheet. Both of these specializations are apparent in star-nosed moles. Although nearly all moles have an array of mechanosensory organs on the nose, the expanded star configuration greatly increases the surface area of the receptors relative to other moles (Fig. 3c, d). The star covers $0.92 cm^2$ per touch, whereas the snout of their close relative, the eastern mole (*Scalopus aquaticus*), covers only $0.11 cm^2$ per touch. Star-nosed moles touch the substrate 13 times per second whereas eastern moles touch eight times per second. Thus, for equivalent small prey densities, star-nosed moles would be expected to encounter roughly 14 times more prey items per unit search time than eastern moles. Figure 2b illustrates the effect of these sensory specializations on R for the two species encountering prey at different densities (assuming equal handling times). However, this comparison remains theoretical because eastern moles do not eat the small prey items eaten by star-nosed moles (see Supplementary Video 4).

As pointed out previously³, at high encounter rates a predator's energy intake (R) approaches a theoretical maximum equivalent to the prey's profitability (E/T_h); that is, T_s approaches zero (arrow Fig. 2c). Star-nosed moles live in wetlands with a high density of invertebrates, and it is likely that they encounter periodic patches of

dense prey. Is it possible for them to achieve the theoretical maximum rate of energy intake of E/T_h when T_s is zero? Our preliminary data from a mole encountering multiple prey items strongly suggest that this is the case. In an astounding flurry of star movements and prey captures, a mole presented with eight separate prey items located and ate them in 1.8 s during the first trial (see Supplementary Video 5). In the second small prey trial, ten separate prey items were located and eaten in 2.3 s. The average handling times for each of the multiple prey items were 229 ms and 233 ms respectively, for each trial—equivalent to the single-prey-item handling times and resulting in a profitability of $43 J s^{-1}$ for these prey; essentially identical to the single-prey-item profitability average of $44 J s^{-1}$. Thus, star-nosed moles can achieve the theoretical

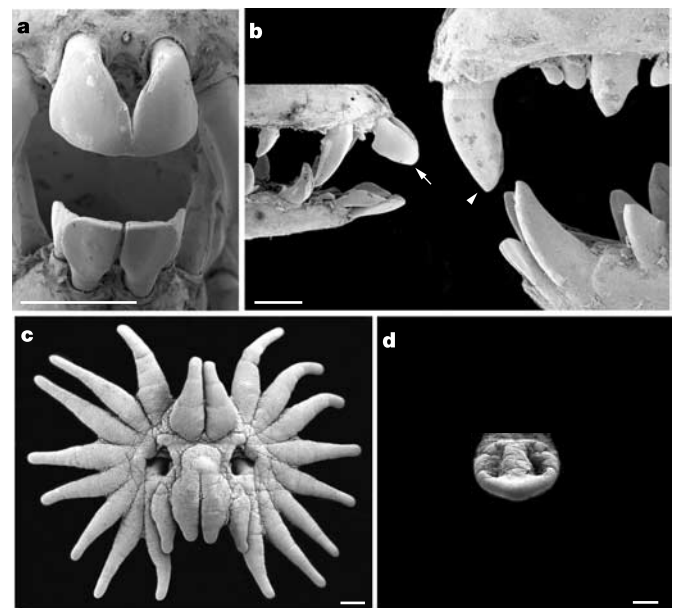


Figure 3 The unusual dentition and nose of star-nosed moles. **a**, **b**, The specialized incisors (arrow) are greatly reduced and modified into a tweezer-like form used to precisely grasp small prey. Compare to eastern mole incisors (arrowhead in **b**). **c**, **d**, The star (**c**) is many times larger than the nose of other moles (**d**), contains twice the density of receptor organs found in other moles, and is divided into a high-resolution central fovea and low-resolution periphery (paralleling trends in high-resolution visual and auditory systems)²⁰. Scale bars, 1 mm.

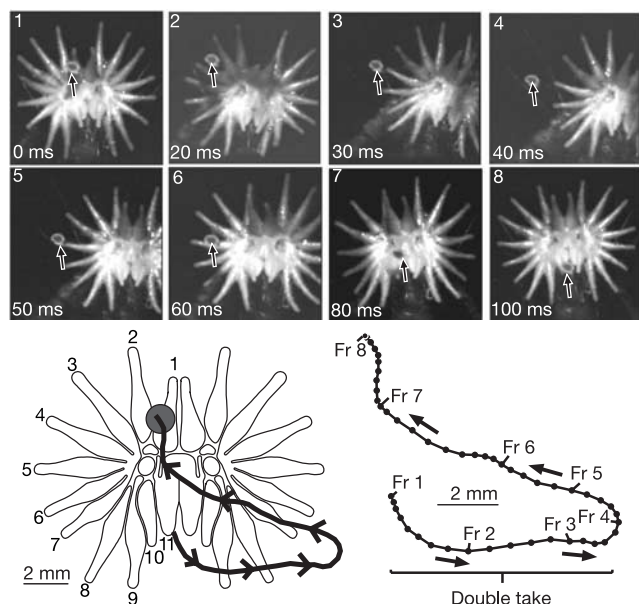


Figure 4 A double take illustrating initial movement in the wrong direction, followed by a sudden correction that brings the foveal region into contact with the prey. Arrows in the top panels (1–8) mark the prey item. The substrate is not contacted during the double take. The lower-left panel shows the track of the nose relative to the prey item (dark sphere). The lower-right panel shows the relative position of the nose every 2 ms (small dots) joined by a line showing the track. Each video frame is indicated (Fr).

maximal R for high densities of small prey—at least in bursts. This may represent the upper limit of energy intake for star-nosed moles consuming the smallest prey items, as details of their behaviour suggest that they have reached the limit of the mammalian sensory system for processing tactile information. For example, star-nosed moles have roughly 25 ms (average touch duration) to make their initial decisions about potential food items before they begin to move the star to the next area of the substrate (Fig. 1a). It takes 12 ms for neurons in the mole's somatosensory cortex to respond to tactile stimulation of the star¹³, and we estimate that it takes roughly 5 ms for a motor command to be conducted to the periphery¹⁴. This leaves little time for central processing and may explain the relatively frequent occurrence of double takes. In over one-third of the trials the star was initially moved in the wrong direction (away from the contacted prey item) before a sudden reversal brought the fovea area into contact with the prey (Fig. 4; see also Supplementary Video 6). This inefficient behaviour lengthens handling times and suggests that star-nosed moles are at, or near, the speed limit for central processing of tactile information.

Supplementary Fig. 1 summarizes our results and their implications for the evolution of sensory-motor systems and foraging strategies. At long handling times there is a low cost to increasing speed because few adaptations are needed to reduce handling time from, for example, 30 to 29 s. However, there is also little to be gained in prey profitability. In contrast, at very short handling times there is a large pay-off in prey profitability in moving from a handling time of 1.5 s to 0.5 s. There are also likely to be great costs in the form of adaptations for speed and, as a corollary, adaptations for detecting smaller prey. Here we suggest that the many sensory-motor specializations characterizing star-nosed moles are the result (and cost) of selective pressure for exploiting high densities of small prey at high speed. The pay-off is not the profitability of individual prey *per se*, but rather the ability to add the resource of small invertebrates, present in the star-nosed mole's wetland habitat⁹, to the diet while maintaining a high average rate of energy intake. But there are limits to how short handling times,

which include pursuit and capture of prey, can become. The ancestor to star-nosed moles probably had characteristics that might be considered exaptations¹⁵ for increasing speed. Specifically, the use of a touch organ near the mouth to detect prey allows for progressive reduction in handling time, whereas distance detection (sight, sound, or smell) limits reduction in handling time because of the need to approach prey. In addition, visual information takes longer to transduce as exemplified by the 200 ms reaction time for a human eye movement¹⁶, longer than the entire handling time for many star-nosed mole trials. Thus, despite potentially high pay-offs, a number of constraints may prevent the evolution of very short handling times in most species. □

Methods

Three star-nosed moles (*Condylura cristata*), two short-tailed shrews (*Blarina brevicauda*) and three eastern moles were collected under the authorization of scientific collecting permits. Moles and shrews were housed separately in 38 × 55 cm containers filled to a depth of 20 cm with moist peat moss. Mole home cages were connected to a separate water container by plastic tubing. The moles' captive diet consisted of commercially raised earthworms. Shrews were fed crickets and mealworms. Small pieces of earthworm were used as prey items for star-nosed moles during behaviour trials. Average size of earthworm pieces for all trials was 1.4 mm, corresponding to a dry weight of approximately 0.5 mg and an energy value of 10 J¹¹. Crickets were used for prey of short-tailed shrews and had an average dry mass of 63 mg, equivalent to approximately 1,400 J¹¹; however, the energy content was reduced by 40% to estimate non-digestible chitin^{17–18}. Behaviour was filmed in a rectangular Plexiglas tunnel system with a removable glass bottom using an S-Series Motionscope (Redlake Imaging Corporation) at a sampling rate of 500 frames per second. Video was stored in a digital buffer and then transferred to VHS tape for scoring. Mole dentition and skin were examined under a Cambridge 360 scanning electron microscope after dehydration and sputter coating with gold. The cortex was fixed in 4% paraformaldehyde, cut tangentially at 60 μm and processed for the metabolic enzyme cytochrome oxidase¹⁹. Mole nose surface area contacts were determined from close-up, videotaped behaviour trials.

Received 18 November; accepted 6 December 2004; doi:10.1038/nature03250.

1. Emlen, J. M. The role of time and energy in food preference. *Am. Nat.* **100**, 611–617 (1966).
2. Schoener, T. W. Theory of feeding strategies. *Annu. Rev. Ecol. Syst.* **2**, 369–404 (1971).
3. Charnov, E. L. Optimal foraging: attack strategy of a mantid. *Am. Nat.* **110**, 141–151 (1976).
4. Pyke, G. H., Pulliam, H. R. & Charnov, E. L. Optimal foraging: a selective review of theory and tests. *Q. Rev. Biol.* **52**, 137–154 (1977).
5. Pyke, G. H. Optimal foraging theory: a critical review. *Annu. Rev. Ecol. Syst.* **15**, 523–575 (1984).
6. Stephens, D. W. & Krebs, J. R. *Foraging Theory* (Princeton Univ. Press, Princeton, 1986).
7. Kamil, A. C. & Sargent, T. D. *Foraging Behavior: Ecological, Ethological, and Psychological Approaches* (Garland STPM, New York, 1981).
8. Catania, K. C. & Kaas, J. H. Somatosensory fovea in the star-nosed mole: behavioral use of the star in relation to innervation patterns and cortical representation. *J. Comp. Neurol.* **387**, 215–233 (1997).
9. Hamilton, W. J. Habits of the star-nosed mole, *Condylura cristata*. *J. Mamm.* **12**, 345–355 (1931).
10. Catania, K. C. & Remple, F. E. Tactile foveation in the star-nosed mole. *Brain Behav. Evol.* **63**, 1–12 (2004).
11. Cummins, K. W. & Wuycheck, J. C. Caloric equivalents for investigations in ecological energetics. *Int. Assoc. Theor. Appl. Limnol.* **18**, 2–158 (1971).
12. Anderson, J. T. & Smith, L. M. Invertebrate response to moist-soil management of playa wetlands. *Ecol. Appl.* **10**, 550–558 (2000).
13. Sachdev, R. N. & Catania, K. C. Receptive fields and response properties of neurons in the star-nosed mole's somatosensory fovea. *J. Neurophysiol.* **87**, 2602–2611 (2002).
14. Maiorov, V. I., Savchenko, E. I. & Kotlyar, B. I. Transformation of the afferent tactile signal into a motor command in the cat motor cortex. *Neurosci. Behav. Physiol.* **10**, 374–381 (1980).
15. Gould, E. & Vrba, E. S. Exaptation: A missing term in the science of form. *Paleobiology* **8**, 4–15 (1982).
16. Carpenter, R. H. S. *Movements of the Eyes* (Pion, London, 1977).
17. Witz, B. W. & Lawrence, J. M. Nutrient absorption efficiencies of the lizard, *Cnemidophorus sexlineatus* (sauria: teiidae). *Comp. Biochem. Physiol.* **105**, 151–155 (1993).
18. Finkler, M. D. Complete nutrient composition of commercially raised invertebrates used as food for insectivores. *Zoo Biology* **21**, 269–285 (2002).
19. Wong-Riley, M. T. & Carroll, E. W. Quantitative light and electron microscopic analysis of cytochrome oxidase-rich zones in V II prestriate cortex of the squirrel monkey. *J. Comp. Neurol.* **222**, 18–37 (1984).
20. Suga, N. & Jen, P. H. Disproportionate tonotopic representation for processing CF-FM sonar signals in the mustache bat auditory cortex. *Science* **194**, 542–544 (1976).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. McCauley, M. Leal, P. Abbot and E. Haldeman for comments on the manuscript. Research was supported by an NSF Career Award to K.C.C.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.C.C. (ken.catania@Vanderbilt.edu).

CD36 is a sensor of diacylglycerides

Kasper Hoebe¹, Philippe Georger¹, Sophie Rutschmann¹, Xin Du¹, Suzanne Mudd¹, Karine Crozat¹, Sosathya Sovath¹, Louis Shamel¹, Thomas Hartung², Ulrich Zähringer³ & Bruce Beutler¹

¹Department of Immunology, The Scripps Research Institute, La Jolla, California 92037, USA

²EU Joint Research Centre, ECVAM, I-21020 Ispra, Italy

³Research Center Borstel, Leibniz-Center for Medicine and Biosciences, 23845 Borstel, Germany

Toll-like receptor 2 (TLR2) is required for the recognition of numerous molecular components of bacteria^{1–8}, fungi^{9,10} and protozoa¹¹. The breadth of the ligand repertoire seems unusual, even if one considers that TLR2 may form heteromers with TLRs 1 and 6 (ref. 12), and it is likely that additional proteins serve as adapters for TLR2 activation. Here we show that an *N*-ethyl-*N*-nitrosourea-induced nonsense mutation of *Cd36* (*oblivious*) causes a recessive immunodeficiency phenotype in which macrophages are insensitive to the *R*-enantiomer of MALP-2 (a diacylated bacterial lipopeptide) and to lipoteichoic acid. Homozygous mice are hypersusceptible to *Staphylococcus aureus* infection. *Cd36^{obl}* macrophages readily detect S-MALP-2, PAM₂CSK₄, PAM₃CSK₄ and zymosan, revealing that some—but not all—TLR2 ligands are dependent on CD36. Already known as a receptor for endogenous molecules, CD36 is also a selective and nonredundant sensor of microbial diacylglycerides that signal via the TLR2/6 heterodimer.

In a G3 population of mice homozygous for *N*-ethyl-*N*-nitrosourea (ENU)-induced mutations¹³, we identified an animal that was hyporesponsive both to macrophage-activating lipopeptide 2 from *Mycoplasma pneumoniae* (MALP-2) and to a crude preparation of commercial peptidoglycan derived from *Staphylococcus aureus*. Biochemical analysis of the peptidoglycan preparation showed that the microbial inducer to which the mutant mouse was insensitive was lipoteichoic acid (LTA; Supplementary Fig. IS-1). Both MALP2 and LTA are TLR2/6-dependent microbial stimuli. However, macrophages from this mouse showed normal responses to the synthetic tri-acylated lipopeptide PAM₃CSK₄ (a TLR1/2 ligand^{5,14}) and zymosan (a TLR2/6 ligand¹²).

This mouse was bred to establish a homozygous stock on a pure C57BL/6 background. The phenotype was named *oblivious* to denote insensitivity to two TLR2 agonists that were both diacylglycerides. *Oblivious* behaved as a strictly recessive mutation (Fig. 1).

The molecular specificity of the sensing deficit in *oblivious* mice was further examined using synthetic *S*- and *R*- enantiomers of MALP-2 as well as diacylated and triacylated molecules with the peptide moiety -Cys-Ser-(Lys)₄ (Supplementary Fig. IS-2a–d). The tumour necrosis factor (TNF)-inducing activities of LTA, S-MALP-2 and R-MALP-2, PAM₂CSK₄ and PAM₃CSK₄ are each fully dependent on TLR2 (Supplementary Fig. IS-2e). However, only responses to R-MALP-2 and LTA were affected by the *oblivious* mutation. In addition, dose–response curves with LTA show that the inhibition observed for *oblivious* macrophages is not complete as observed for TLR2 knockout mice, suggesting that the *oblivious* gene product facilitates but is not absolutely required for LTA sensing (Supplementary Fig. IS-2f). Consistent with this finding was our observation of partial preservation of MyD88-dependent signal transduction events when either LTA or MALP-2 was used as an inducing stimulus in *oblivious* macrophages (Supplementary Fig. IS-3). It may be concluded that *oblivious*-assisted activation of TLR2 depends on the number of fatty acid groups that are present, the chirality, and the size and/or composition of the peptide group.

Oblivious homozygotes all develop an ocular pterygium by 6–12 months of age, apparently as the result of spontaneous colonization with *Corynebacterium* species and *Staphylococcus lentus*. On occasion, colonization progresses to spontaneous endophthalmitis (Supplementary Fig. IS-4). Luminescent *Staphylococcus aureus* was used as a direct test of the ability of *oblivious* mice to control an intradermal and/or intravenous Gram-positive inoculum. *Tlr2*-null and/or *Tirap*-null mice were taken as controls to assess the severity of the phenotype. During a 7-day period, *oblivious* and *Tlr2*^{−/−} mice were permissive for the growth of intradermally inoculated *S. aureus*, whereas the bacteria were cleared in wild-type C57BL/6 mice (Fig. 2a). Survival of wild-type, *oblivious*, *Tlr2*^{−/−} and *Tirap*^{−/−} mice and bacterial growth was monitored during a 10-day period after challenge with 10⁸ colony-forming units (c.f.u.) of *S. aureus* per mouse, administered intravenously. *Tlr2*^{−/−} mice all succumbed to infection within 5 days. *Tirap*^{−/−} and *oblivious* mice were less susceptible but accumulated more organisms during the course of infection, and died more frequently than wild-type mice (Fig. 2b, c). This is consistent with analyses of TNF production *in vitro* by peritoneal macrophages stimulated with LTA or MALP-2. *Tirap*^{−/−} and *oblivious* cells showed a residual response, whereas *Tlr2*^{−/−} cells were completely unresponsive (Fig. 2d). *In vitro*

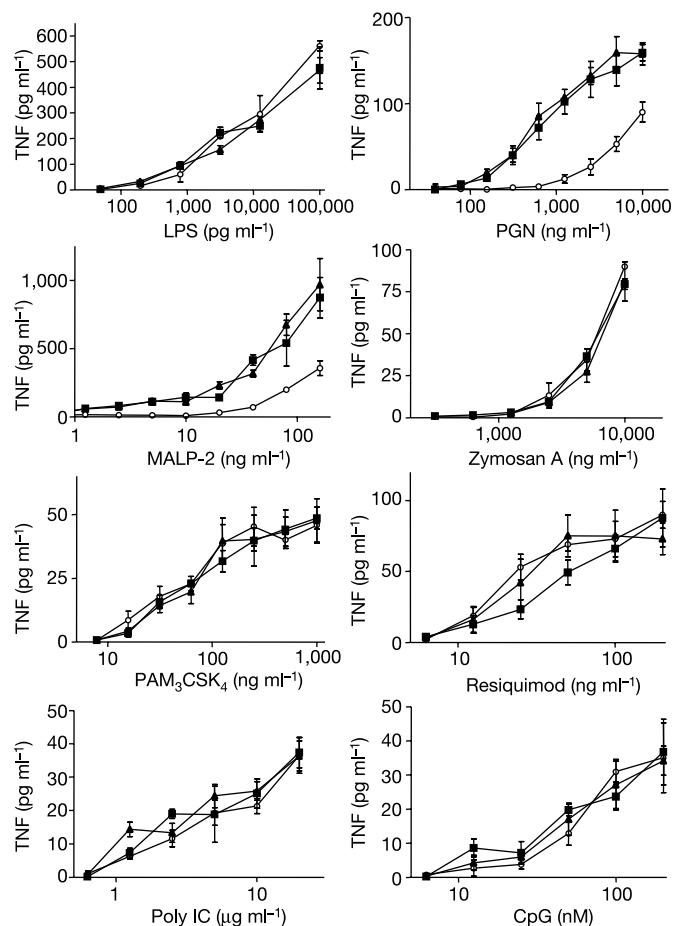


Figure 1 *Oblivious* specifically reduces TNF production in response to *Staphylococcus*-derived peptidoglycan and synthetic MALP-2. Mice were injected with 3% thioglycollate; after 3 days, macrophages were cultured at a density of 5×10^5 cells per well in 96-well plates. Wild-type (solid squares), *obl/+* (solid triangle) and *obl/obl* (open circle) macrophages were incubated with various concentrations of each specific inducer. After 4 h of incubation, supernatants were collected and assayed in duplicate for TNF activity as described previously²⁴. Values are means $\pm$ s.e.m. ($n = 4$ mice).

responses to heat-killed *S. aureus* were similar, although the differences were less pronounced than observed for LTA or MALP2 (Fig. 2e).

The *oblivious* phenotype was mapped by outcrossing homozygotes to C3H/HeN mice and backcrossing the progeny to the mutant stock. On the basis of 28 meioses genotyped at 59 genome-wide microsatellite markers, *obl* was assigned to chromosome 5 (Fig. 3a). Further mapping on a total of 658 meioses permitted confinement to a 1.15-megabase chromosomal interval containing five genes (Fig. 3b, Supplementary Table 1). Sequencing revealed a point mutation (nucleotide 1022 T → A with respect to GenBank accession no. NM_007643) in the distal coding region of *Cd36* (exon 11), creating an early stop codon and removing 133 residues from the carboxy terminus of the polypeptide chain, thereby preventing the formation of the second of two transmembrane domains represented in the normal protein (Fig. 3c–e). No mutations were identified within the other genes of the critical region, all of which were bidirectionally sequenced at complementary DNA or genomic levels.

We analysed CD36 expression on *+/+*, *obl/+* and *obl/obl* macrophages using flow cytometry. *obl/obl* macrophages and platelets showed no expression of CD36 on the surface, whereas *obl/+* macrophages showed reduced expression compared with *+/+* macrophages (Fig. 4a, b). Western blot analysis with antibodies

directed against the amino-terminal part of CD36 showed an absence of CD36 in whole cell extracts as well as in supernatant (Supplementary Fig. IS-5). The phenotype observed for *CD36^{obl/obl}* was identical to that observed in *CD36^{-/-}* mice (Fig. 4c). Finally, *obl/obl* cells that were transfected with an expression vector encoding normal CD36 were capable of producing TNF in response to LTA (Fig. 4d). We therefore concluded that the *Cd36^{obl}* mutation was solely responsible for the *oblivious* phenotype and that the *Cd36^{obl}* allele is functionally equivalent to a null allele.

HEK-293 cells were transfected to express CD36, in combination with TLR2 and/or TLR6 and a nuclear factor (NF)-κB reporter gene. LTA and MALP-2 both caused NF-κB activation when cells were transfected to express TLR2, whereas transfection with plasmids encoding CD36 or TLR6 alone did not support activation by either ligand. The response to LTA and S-MALP was significantly augmented when either CD36 or TLR6 was coexpressed with TLR2. Coexpression of TLR2, TLR6 and CD36 gave the highest responses of all, indicating that CD36 does indeed facilitate the response to LTA and S-MALP, and does so best in the presence of TLR6, although neither TLR6 nor CD36 is absolutely required for a response (Supplementary Fig. IS-6).

CD36 is a member of the scavenger receptor type B family (Supplementary Fig. IS-7)¹⁵. It has been implicated in the recognition of oxidized LDL particles and the uptake of fatty acids^{16,17}.

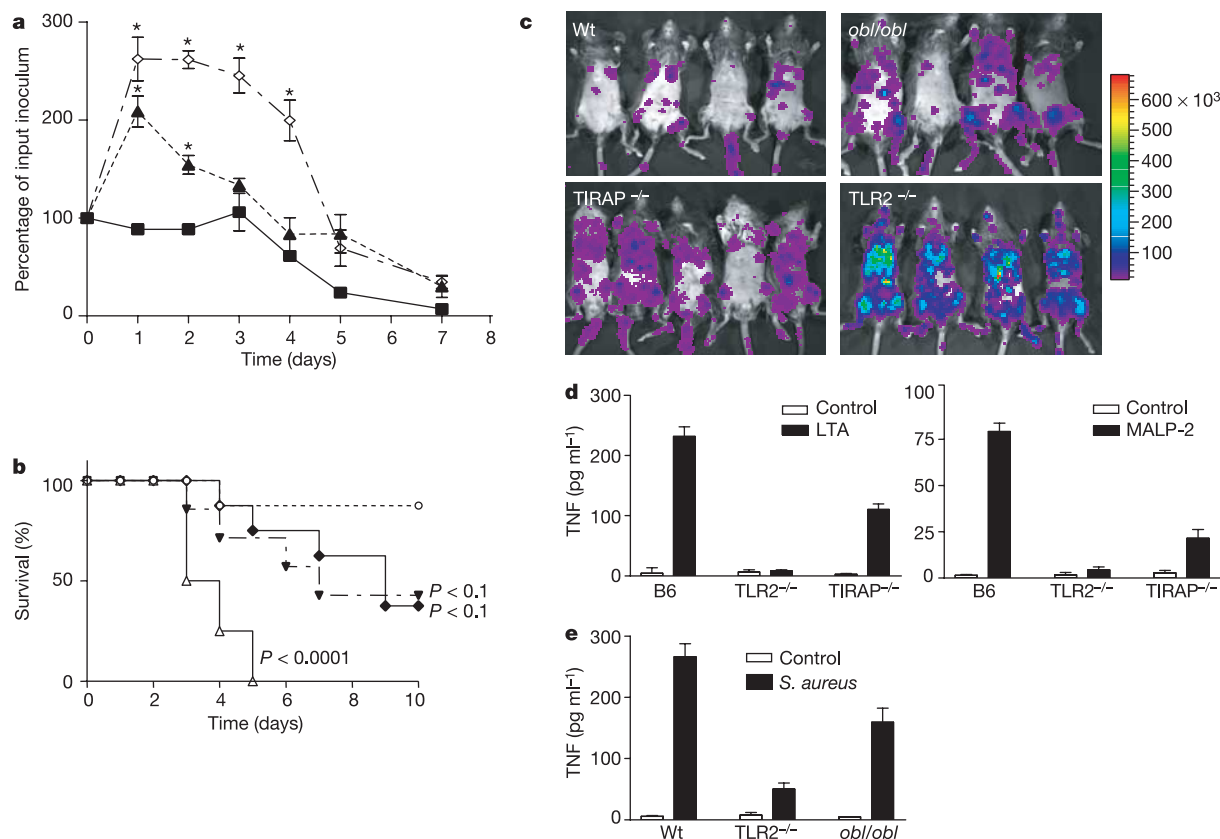


Figure 2 *Oblivious* homozygous mutant mice are unable to contain Gram-positive *S. aureus* infections. **a**, Wild-type C57BL/6 male mice (filled squares), TLR2 knockout (open diamonds) and homozygous *oblivious* (filled triangles) mutant mice were each injected subcutaneously with *S. aureus* (10^5 c.f.u. ml⁻¹). During a 7-day period, bacterial growth was monitored daily with the Xenogen IVIS imaging system. Luminescence is depicted as a percentage of input, with 100% being the luminescence at day zero. Data are shown as means ± s.e.m. (*n* = 6). Asterisk, *P* < 0.001 (compared with wild-type mice). **b**, Survival of wild-type C57BL/6 mice (open circles), *Tlr2*^{-/-} (open triangles), *Tlr2*^{-/-} (filled triangles) and *obl/obl* (filled diamonds) mutant mice after intravenous

administration of *S. aureus* (10^8 c.f.u. per mouse); *n* = 8 for each group; *P* refers to comparison with C57BL/6. **c**, Luminescence in normal C57BL/6 (Wt), *Tlr2*^{-/-}, *Tlr2*^{-/-} and *obl/obl* mutant mice at day 2 after intravenous injection with *S. aureus* (10^8 c.f.u. ml⁻¹). **d**, TNF production induced by LTA (left) and MALP-2 (right) in peritoneal macrophages derived from wild-type, *Tlr2*^{-/-} and *Tlr2*^{-/-} mice during a 4-h incubation. **e**, TNF production induced by heat-killed *S. aureus* (10^6 c.f.u. ml⁻¹) in peritoneal macrophages derived from wild-type, *Tlr2*^{-/-} and *obl/obl* mice during a 4-h incubation. Error bars represent s.e.m. of three independent experiments.

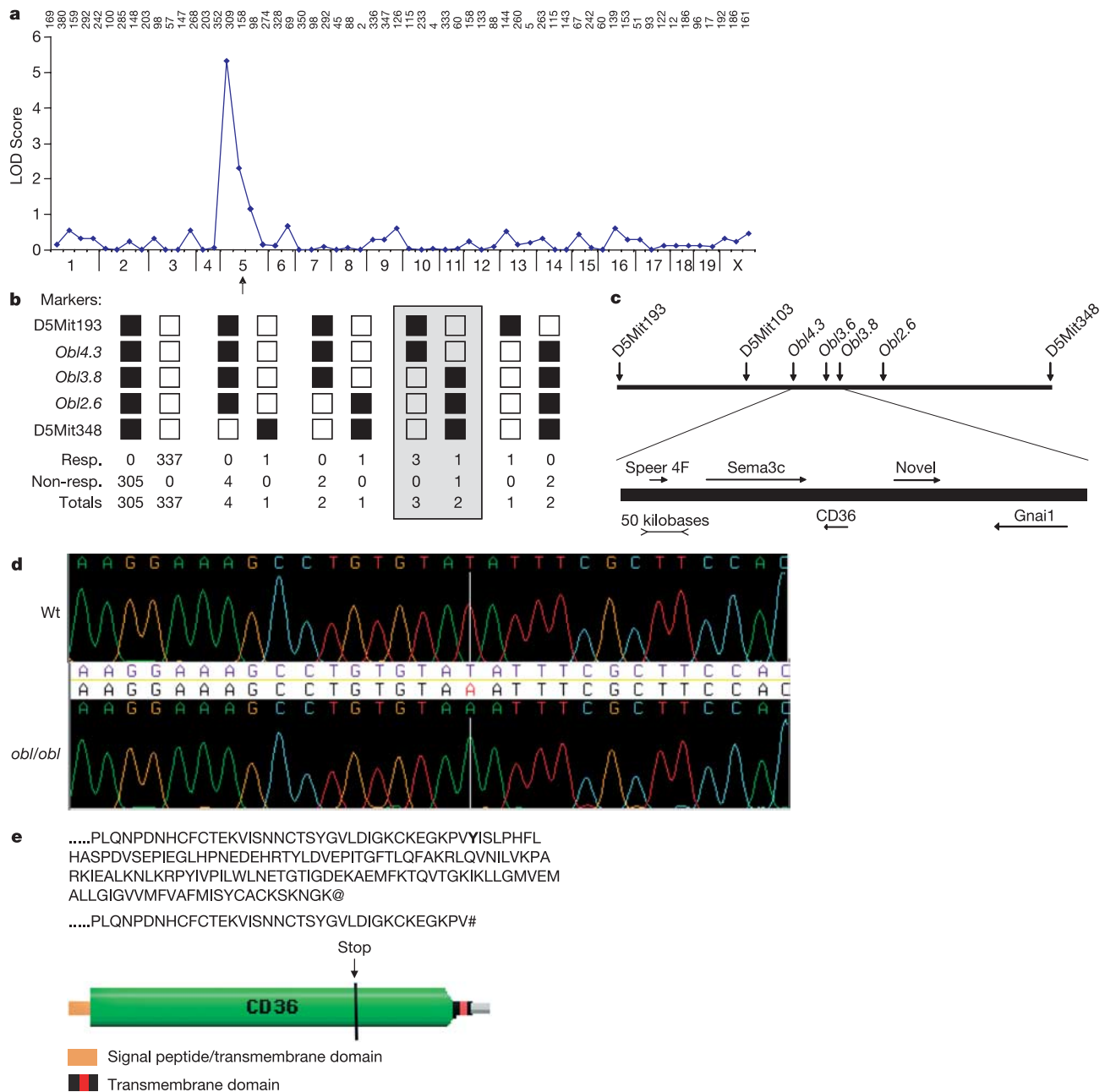


Figure 3 Confinement of the *ob/* critical region and identification of the *ob/* mutation. **a**, Coarse genetic mapping of *oblivious*. Phenotypic classification of F_2 mice was based on a comparison of peptidoglycan-induced dose-response curves by measuring TNF production with the L929 bioassay. On 29 meioses, *ob/* was confined to the proximal half of chromosome 5 with a peak LOD score of 5.3 associated with D5Mit352. **b**, On 658 meioses, the mutation was confined between novel polymorphic microsatellites *Obi4.3* and *Obi3.8* (shaded box; markers shown in Supplementary Table 1), and is separated from the limiting markers by four crossovers on the proximal site and one confirmed crossover on the distal side. Resp., 'responder' (to peptidoglycan); non-resp., 'non-responder'; filled squares, homozygous; open squares, heterozygous. **c**, The critical

region is flanked by the public markers D5Mit193 and D5Mit348, and by the informative microsatellites shown in Supplementary Table 1. Within the critical region, a nominal listing of five genes is presented in the Ensembl database, with the arrangement and orientation shown. **d**, Consed display (<http://www.phrap.org/consed/consed.html>) of the mutation, showing the sequence from the distal coding region of a *Cd36^{ob/ob}* homozygote (bottom trace) and from a normal C57BL/6 mouse (top trace). The mutation involves a single base transversion from T to A in the mutant strain, causing the nonsense substitution Y340#. **e**, The predicted sequence and graphical display of the Cd36 protein in normal (top) and mutant (bottom) mice.

However, no defined molecules of eubacterial origin have previously been shown to signal through CD36, nor has any function of CD36 previously been shown to be dependent on TLR. Our data suggest that CD36 acts as a facilitator or co-receptor for di-acylglyceride recognition through the TLR2/6 complex. Thus, CD36 serves a function analogous to CD14, which concentrates the lipopolysaccharide (LPS) signal for trans-

duction through TLR4 (ref. 18).

CD36 also has an established role in the recognition of endogenous ligands¹⁹ and might therefore represent a signalling bridge between endogenous inducers and the principal innate immune sensors of the mammalian host. Potentially, such a bridge might be important in the pathogenesis of sterile inflammatory conditions.

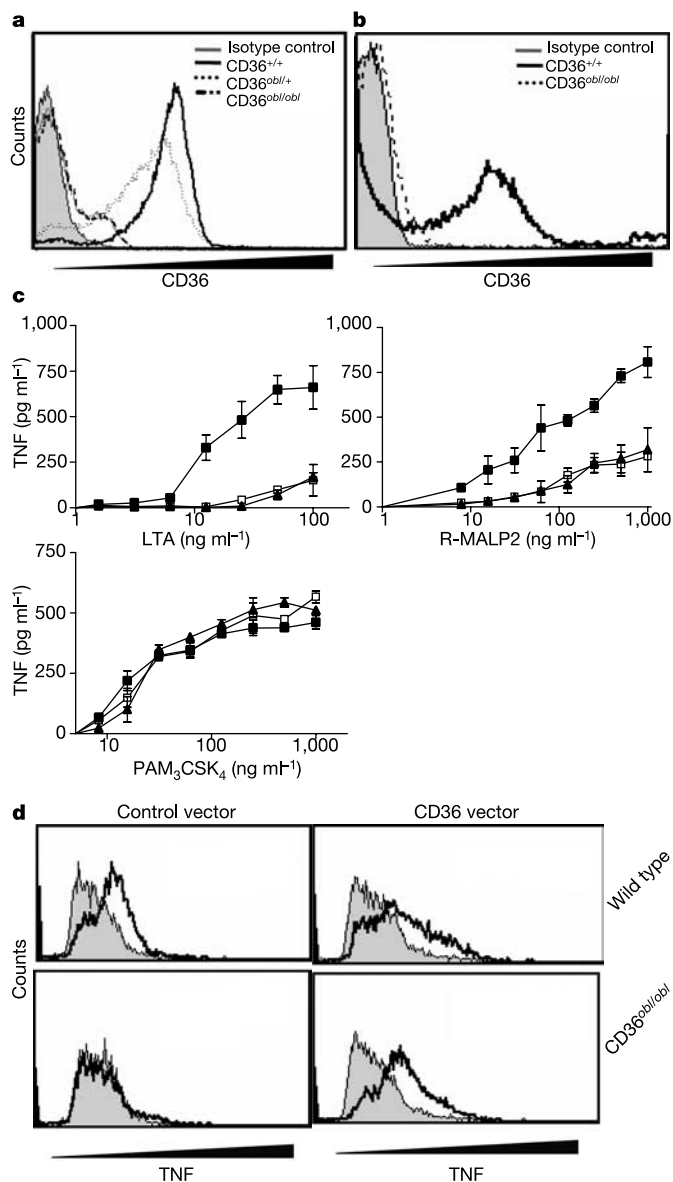


Figure 4 A nonsense mutation in CD36 is responsible for the *ob/ob* phenotype. **a, b**, Surface expression of CD36 on peritoneal macrophages (**a**) and platelets (**b**) derived from wild-type, *obl/+* and/or *obl/obl* mice. Cells were labelled with a primary monoclonal antibody against mouse CD36 or with an isotype control, then exposed to a secondary anti-mouse IgA coupled to phycoerythrin. **c**, Comparison of peritoneal macrophage responses between wild-type (filled squares), *Cd36*^{-/-} (filled triangles) and *Cd36*^{obl/obl} (open squares) mice when exposed to LTA, R-MALP2 or PAM₃CSK₄. Error bars represent s.e.m. of three independent experiments. **d**, Rescue of the phenotype after co-transfection of bone-marrow-derived macrophages from normal or homozygous mutant mice with an expression vector encoding yellow fluorescent protein as well as an empty vector (control), or with a 1:3 mixture of expression vectors encoding yellow fluorescent protein and normal Cd36 as described before²³. Untreated, shaded histogram; LTA (10 ng ml⁻¹)-treated, unshaded histogram.

Methods

Mice

TLR2 knockout mice were provided by Tularik Inc., and Tirap knockout mice were obtained from R. Medzhitov (Yale University School of Medicine, New Haven, Connecticut). CD36 knockout mice were a gift from M. Febbraio (Cleveland Clinic Foundation, Cleveland, Ohio).

Innate immune activators, probes and antibodies

Salmonella minnesota Re595 LPS was obtained from Alexis. dsRNA was purchased from Amersham Pharmacia Biotech. Racemic MALP-2, S-MALP-2 and R-MALP-2, PAM₃CSK₄

and PAM₂CSK₄ were obtained from EMC microcollections GmbH and peptidoglycan was purchased from Fluka. Phosphorothioate-stabilized CpG oligodeoxynucleotide (CpG-ODN) 5'-TCC-ATG-ACG-TTC-CTG-ATG-CT-3' was obtained from Integrated DNA Technologies. Phospho-specific antibodies for extracellular signal-related kinases (ERKs), p38, c-Jun N-terminal kinase (JNK) and antibodies against IκB were obtained from Cell Signalling Technology. Monoclonal antibodies against mouse CD36 (MAB1258) were obtained from Chemicon International and polyclonal antibodies against human CD36 were obtained from Cayman Chemical. Phycoerythrin-labelled anti-mouse IgA was purchased from E-Bioscience. Purified LTA was prepared as described previously⁶.

Purification and analysis of the bioactive material in peptidoglycan

Insoluble peptidoglycan (uPG) was prepared from *Staphylococcus aureus* and residual teichoic acid was removed as described²⁰. Chloroform/methanol extraction of this uPG preparation revealed bioactivity in the organic phase, strongly implicating lipids rather than PG as the CD36 ligand(s). Further purification of these lipids by silica-gel chromatography and subsequent analysis of bioactive fractions by mass spectrometry revealed several lipophilic partial structures of LTA, including gentiobioside-diaclylglycerol (the LTA lipid anchor)²¹ carrying one glycerol phosphate unit (Gro-P-β-D-Glcp-(1 → 6)-β-D-Glcp-DAG, where Gro stands for glycerol, Glc for glucosyl, p for phospho, and DAG for diacylglycerol).

Western blotting

Western blot analysis of JNK, p38, STAT and ERK1/2 phosphorylation, and degradation of IκB was performed as described previously²². For western blot analysis of CD36 in cell extracts or supernatants, anti-(human CD36) polyclonal antibodies showing crossreactivity to mouse CD36 were used. For analysis of supernatants, macrophages were incubated for 24 h in a small volume of medium and without the addition of fetal bovine serum.

Rescue studies

Rescue studies with bone marrow cells were performed as described previously²³.

Transfection of HEK-293 cells

Cells were cultured in 96-well plates at a density of 2.5×10^4 cells per well and incubated overnight at 37 °C and 5% CO₂. Cells were transfected overnight with NF-κB luciferase reporter gene (50 ng per well) and co-transfected with CD36, TLR2 and/or TLR6 (100 ng per vector per well; final concentration 350 ng per well). After stimulation for 5 h with various concentrations of LTA, S-MALP and PAM₃CSK₄, cells were lysed and luciferase activity was measured with a luciferase assay reagent (Promega).

Germline mutagenesis

ENU was obtained from Sigma. ENU germline mutagenesis was performed as described previously²³. After identification of the *oblivious* mutant in the G3 macrophage screen, the founder was backcrossed to C57BL/6 mice to obtain further genetic purity.

DNA sequencing

Genes residing in the critical region were amplified by using primers chosen after masking with RepeatMasker. Fragments were sequenced bidirectionally with internal primers. Primers were obtained from Genosys.

Infection with *Staphylococcus aureus*

Hair was removed by chemical depilation and the mice were injected subcutaneously with 100 μl of a mid-exponential growth phase ($D_{600} = 0.6$, 5×10^6 c.f.u. ml⁻¹) of *S. aureus* (Xen 8.1; parental strain 3825-4) resuspended in PBS and complexed to Cytodex beads (Sigma) as a carrier. During a 4-day period, bacterial growth was monitored daily with the Xenogen IVIS imaging system. In addition, in a separate experiment mice were administered 10^8 *S. aureus* intravenously without the use of Cytodex beads and were monitored for survival and bacterial growth during a 10-day period.

Received 1 November; accepted 3 December 2004; doi:10.1038/nature03253.

- Lien, E. *et al.* Toll-like receptor 2 functions as a pattern recognition receptor for diverse bacterial products. *J. Biol. Chem.* **274**, 33419–33425 (1999).
- Takeuchi, O. *et al.* Differential roles of TLR2 and TLR4 in recognition of Gram-negative and Gram-positive bacterial cell wall components. *Immunity* **11**, 443–451 (1999).
- Means, T. K. *et al.* Human toll-like receptors mediate cellular activation by *Mycobacterium tuberculosis*. *J. Immunol.* **163**, 3920–3927 (1999).
- Takeuchi, O. *et al.* Discrimination of bacterial lipoproteins by Toll-like receptor 6. *Int. Immunol.* **13**, 933–940 (2001).
- Takeuchi, O. *et al.* Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *J. Immunol.* **169**, 10–14 (2002).
- Morath, S., Stadelmaier, A., Geyer, A., Schmidt, R. R. & Hartung, T. Synthetic lipoteichoic acid from *Staphylococcus aureus* is a potent stimulus of cytokine release. *J. Exp. Med.* **195**, 1635–1640 (2002).
- Opitz, B. *et al.* Toll-like receptor-2 mediates *Treponema glycolipid* and lipoteichoic acid-induced NF-κB translocation. *J. Biol. Chem.* **276**, 22041–22047 (2001).
- Werts, C. *et al.* Leptospiral lipopolysaccharide activates cells through a TLR2-dependent mechanism. *Nature Immunol.* **2**, 346–352 (2001).
- Gantner, B. N., Simmons, R. M., Canavera, S. J., Akira, S. & Underhill, D. M. Collaborative induction of inflammatory responses by dectin-1 and Toll-like receptor 2. *J. Exp. Med.* **197**, 1107–1117 (2003).
- Brown, G. D. *et al.* Dectin-1 mediates the biological effects of β-glucans. *J. Exp. Med.* **197**, 1119–1124 (2003).
- Campos, M. A. *et al.* Activation of Toll-like receptor-2 by glycosylphosphatidylinositol anchors from a protozoan parasite. *J. Immunol.* **167**, 416–423 (2001).
- Ozinsky, A. *et al.* The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between Toll-like receptors. *Proc. Natl. Acad. Sci. USA* **97**, 13766–13771 (2000).
- Hoeb, K. *et al.* Identification of Lps2 as a key transducer of MyD88-independent TIR signaling.

Nature 424, 743–748 (2003).

14. Morr, M., Takeuchi, O., Akira, S., Simon, M. M. & Muhlrad, P. F. Differential recognition of structural details of bacterial lipopeptides by Toll-like receptors. *Eur. J. Immunol.* **32**, 3337–3347 (2002).
15. Calvo, D., Dopazo, J. & Vega, M. A. The CD36, CLA-1 (CD36L1), and LIMPII (CD36L2) gene family: cellular distribution, chromosomal location, and genetic evolution. *Genomics* **25**, 100–106 (1995).
16. Aitman, T. J. *et al.* Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genet.* **21**, 76–83 (1999).
17. Glazier, A. M., Scott, J. & Aitman, T. J. Molecular basis of the Cd36 chromosomal deletion underlying SHR defects in insulin action and fatty acid metabolism. *Mamm. Genome* **13**, 108–113 (2002).
18. Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J. & Mathison, J. C. CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science* **249**, 1431–1433 (1990).
19. Gordon, S. Pattern recognition receptors: doubling up for the innate immune response. *Cell* **111**, 927–930 (2002).
20. Rosenthal, R. S. & Dziarski, R. Isolation of peptidoglycan and soluble peptidoglycan fragments. *Methods Enzymol.* **235**, 253–285 (1994).
21. Deininger, S. *et al.* Definition of structural prerequisites for lipoteichoic acid-inducible cytokine induction by synthetic derivatives. *J. Immunol.* **170**, 4134–4138 (2003).
22. Kim, S. O., Ono, K. & Han, J. Apoptosis by pan-caspase inhibitors in lipopolysaccharide-activated macrophages. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **281**, L1095–L1105 (2001).
23. Hoebe, K., Du, X., Goode, J., Mann, N. & Beutler, B. *Lps2*: a new locus required for responses to lipopolysaccharide, revealed by germline mutagenesis and phenotypic screening. *J. Endotoxin Res.* **9**, 250–255 (2003).
24. Poltorak, A. *et al.* Genetic and physical mapping of the *Lps* locus-identification of the toll-4 receptor as a candidate gene in the critical region. *Blood Cells Mol. Dis.* **24**, 340–355 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.B. (bruce@scripps.edu).

Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots

Haruko Imaizumi-Anraku^{1,2*}, Naoya Takeda^{3*}, Myriam Charpentier^{4†}, Jillian Perry⁴, Hiroki Miwa⁵, Yosuke Umehara^{1,6}, Hiroshi Kouchi^{1,6}, Yasuhiro Murakami^{1,2}, Lonneke Mulder⁴, Kate Vickers⁴, Jodie Pike⁴, J. Allan Downie⁵, Trevor Wang⁵, Shusei Sato⁷, Erika Asamizu⁷, Satoshi Tabata⁷, Makoto Yoshikawa³, Yoshikatsu Murooka³, Guo-Jiang Wu⁸, Masayoshi Kawaguchi^{6,8}, Shinji Kawasaki^{1,2}, Martin Parniske^{4†} & Makoto Hayashi^{3,6}

¹National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan

²Bio-oriented Technology Research Advancement Institution (BRAIN), Tokyo Office, 3-18-19 Toranomon Minato-ku, Tokyo 105-0001, Japan

³Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

⁴The Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

⁵John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

⁶Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0112, Japan

⁷Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan

⁸Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan

* These authors contributed equally to this work

† Present address: Genetics Institute, Ludwig Maximilians Universität (LMU), Maria-Ward-Str. 1a, D-80638 München, Germany

The roots of most higher plants form arbuscular mycorrhiza, an ancient, phosphate-acquiring symbiosis with fungi, whereas only four related plant orders are able to engage in the evolutionary younger nitrogen-fixing root-nodule symbiosis with bacteria¹. Plant symbioses with bacteria and fungi require a set of common

signal transduction components that redirect root cell development^{2,3}. Here we present two highly homologous genes from *Lotus japonicus*, *CASTOR* and *POLLUX*, that are indispensable for microbial admission into plant cells and act upstream of intracellular calcium spiking⁴, one of the earliest plant responses to symbiotic stimulation. Surprisingly, both twin proteins are localized in the plastids of root cells, indicating a previously unrecognized role of this ancient endosymbiont in controlling intracellular symbioses that evolved more recently.

Upon recognition of a bacterial symbiotic signalling molecule, the ‘Nod factor’, root hairs of leguminous plants reprogramme their development^{5,6}. Tip growth is redirected to form a curl that entraps a bacterial microcolony to initiate the formation of the infection thread. These events are preceded by the rapid formation of transient subcellular gradients of proton, potassium, chloride and calcium ions, followed after about 10 min by oscillations in intracellular calcium concentration, the ‘calcium spiking’ response⁷. We used *Lotus japonicus* mutants to identify the molecular players of this electrochemical prelude to symbiosis. Mutants affected in the *CASTOR* or *POLLUX* genes exhibit very similar phenotypes; they do not form root nodules or arbuscular mycorrhizal symbiosis. Root hairs treated with *Mesorhizobium loti* or Nod factor show branching and tip swelling (Fig. 1a–f, and data not shown), but infection threads are never formed. Similarly, in the arbuscular mycorrhizal symbiosis, fungal entry into root epidermal cells is not supported, and infection attempts are aborted^{8–11}. Root hairs of all tested *castor*¹² and *pollux* mutants lacked the Nod-factor-induced calcium spiking that is typically seen with wild-type (Fig. 1g). However, root-hair branching was observed in these non-spiking mutants (Fig. 1e, f) and so some of the earliest plant responses to Nod factors can be genetically uncoupled. These results demonstrate that *CASTOR* and *POLLUX* act very early in a signal transduction chain leading from the perception of Nod factor to the activation of calcium spiking. They are also required for the execution of the appropriate morphological response of the root hair, because only branching and deformation, but not curling or infection thread formation, were observed.

The *CASTOR* and *POLLUX* genes were isolated through positional cloning and by their strong sequence similarity. Using the extensive collection of microsatellite (simple sequence repeat (SSR)) markers available for *L. japonicus*¹³, the *CASTOR* and *POLLUX* genes were positioned close to the bottom ends of chromosomes 1 and 6, respectively. Analysis of the sequence obtained in the context of the genome project¹⁴ revealed very limited local similarity between the *CASTOR* and *POLLUX* genomic regions. Only *CASTOR* and *POLLUX* themselves were conserved, suggesting that the two genes are the product of a limited duplication event (Supplementary Fig. 1).

Five *castor* alleles were independently mapped to the south end of chromosome 1, close to SSR marker TM0105 (ref. 14). Starting with this marker, a bacterial artificial chromosome (BAC)¹⁵/transformation-competent artificial chromosome (TAC)¹⁴ contig (Supplementary Fig. 1b) was constructed that extended into the telomeric region, as indicated by multiple copies of subtelomeric repeat LjTR4 (M.H., unpublished data) towards the southern extremity (Supplementary Fig. 1b). Genetic mapping was affected by an inversion 145 kilobases (kb) long between the parental lines B-129 and MG-20 (ref. 16) with no recombination events observed on inspection of 1,833 individuals of five different F₂ populations (Supplementary Fig. 1a), delimiting the *CASTOR* locus to a region of about 240 kb. Polymerase chain reaction (PCR) marker TB2R could not be amplified from individuals homozygous for the *castor-8* or *castor-9* alleles. PCR and Southern blot analysis (Fig. 2a) indicated that both alleles have large deletions of more than 20 kb encompassing the *CASTOR* gene, and subsequent sequence analysis revealed that marker TB2R is located in a *CASTOR* intron (Supplementary Fig. 1d).

Sequencing of genomic DNA identified non-silent mutations within the *CASTOR* gene in 15 additional mutants, thus providing a

proof of gene identification (Supplementary Table1). Four independent mutations (*castor-8* to *castor-11*), which arose during tissue culture and are thus attributable to somaclonal variation, were deletions between 1 base pair and more than 20 kb in size (Supplementary Table 1), and not retrotransposon insertions as observed in other *Lotus* somaclonal mutants¹⁷. This bias might be caused by the proximity to the telomere.

The *POLLUX* gene was found to segregate with SSR marker TM0885 (Supplementary Fig. 1e). A TAC clone carrying SSR

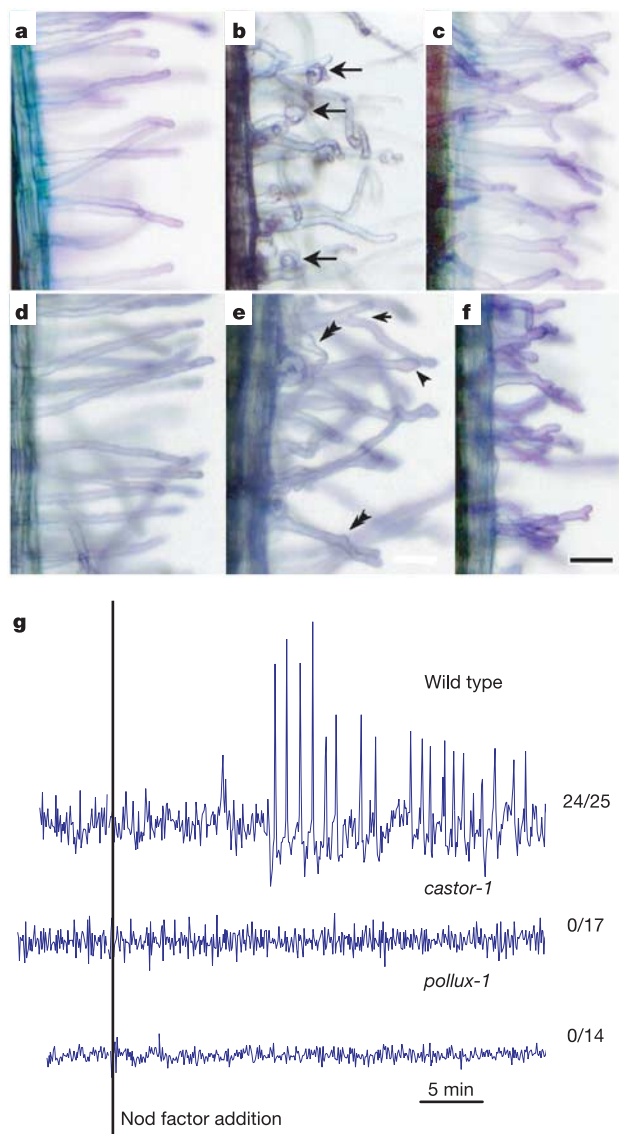


Figure 1 Phenotypes of *castor* and *pollux* mutants. **a–f**, *L. japonicus* root hair responses induced by *Mesorhizobium loti* TONO inoculation or Nod factor (NF) treatment. Root hairs of wild-type (B-129 Gifu) and *castor* mutants, inoculated with *M. loti* TONO (**b**, **e**) or treatment with 100 nM NF (**c**, **f**), are shown. **a**, **d**, Non-treated control of wild type and *castor-4*, respectively. **b**, Wild-type root hair curling (arrows). **e**, Root hair swelling (arrowhead), tip growth (double arrowheads) and branching (small arrow) on a *castor-5* mutant inoculated with *M. loti*. **c**, **f**, NF-induced root hair deformation, on wild-type and *castor-4* mutant roots, respectively. Scale bar, 50 μ m. **g**, Analysis of calcium spiking in root hairs. Each trace is from a single root hair using seedlings of the wild type (B-129 Gifu) or of mutants carrying the *castor-1* or *pollux-1* alleles. Root hairs were injected with the calcium-sensitive dye Oregon green-dextran and after about 20 min NF was added at 10 nM. The graph shows typical traces of the differences in fluorescence intensity between 5-s sequential time points. Similar results were seen with the *castor-2*, *castor-3*, *castor-4* and *pollux-2* mutants; the fractions show the number of root hairs inducing calcium spiking relative to the numbers assayed for each locus, with separate plants for each assay.

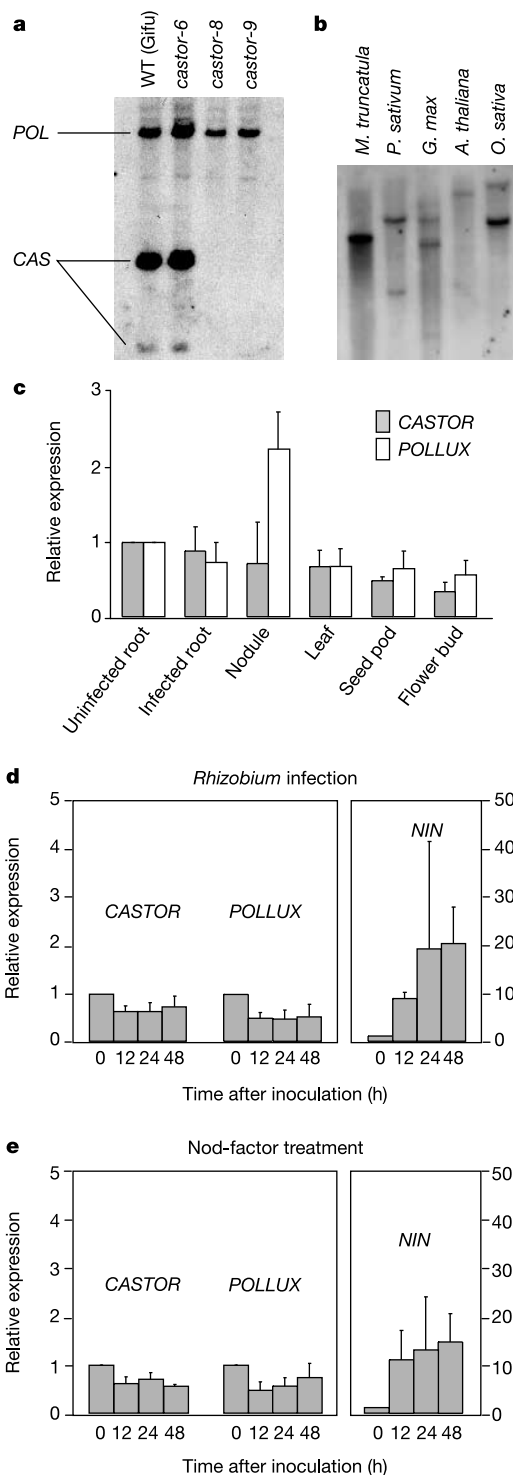


Figure 2 Southern blot and RT-PCR expression analyses of *CASTOR* and *POLLUX*. **a**, **b**, DNA gel blot analysis of *CASTOR* (CAS) and *POLLUX* (POL). **a**, Wild-type (WT; B-129 Gifu) and *castor* mutant alleles (*castor-6*, *castor-8* and *castor-9*). Two distinct signals, representing the CAS and POL genes, were detected with a *CASTOR* probe on DNA of the wild type or of plants homozygous for point-mutation allele *castor-6*. The fragment representing CAS was absent from the DNA of deletion mutants *castor-8* and *castor-9*. **b**, Leguminous and non-leguminous plants. *G. max*, *Glycine max*; *O. sativa*, *Oryza sativa*. Genomic DNAs were digested with *Xba*I (**a**) or *Bam*HI (**b**). **c**, *CASTOR* and *POLLUX* expression in various organs of *L. japonicus* B-129 Gifu. **d**, **e**, Induction of *CASTOR*, *POLLUX* and *NIN* expression from 0 to 48 h after inoculation with *M. loti* (**d**) or treatment with Nod factor (**e**). Expression strengths are shown relative to uninfected root (**c**, **d**) or non-treated root (**e**). The error bars indicate standard deviation.

TM0885 was completely sequenced, and a candidate gene with high homology to *CASTOR* was identified. Sequencing of nine different symbiosis-defective mutant alleles confirmed the identification of the *POLLUX* gene (Supplementary Table 1).

Full-length complementary DNA (cDNA) sequences corresponding to the *CASTOR* and *POLLUX* genes were obtained through 5' and 3' rapid amplification of cDNA ends and by sequencing of cDNA clones. Alignment of the cDNA with the genomic sequence identified 12 exons for both genes but also the occurrence of alternative splicing (Supplementary Fig. 1d, f). Interestingly, the intron between the seventh and eighth exon was retained in several *CASTOR* cDNA clones, and the presence of this non-spliced intron was also confirmed by reverse transcriptase PCR (RT-PCR) on root and nodule RNA (data not shown). This alternatively spliced mRNA of unknown function encodes a truncated protein of 569 amino acids.

The *POLLUX* gene was detected by genomic DNA-blot analysis using *CASTOR* as a probe (Fig. 2a). *CASTOR*–*POLLUX* homologues exist in one or two copies in other dicotyledonous and monocotyledonous plants (Fig. 2b). Notably, only a single signal was detected in *Medicago truncatula*. This signal probably corresponds to the recently cloned *DMI1* gene¹⁸, whose product is more closely related to *POLLUX* than it is to *CASTOR* (Fig. 3d).

The expression level of *CASTOR* and *POLLUX* mRNA in each

organ and in early stages of symbiosis was examined by quantitative RT-PCR (Fig. 2c–e). *CASTOR* was expressed in all organs, with highest expression in non-infected roots, whereas *POLLUX* expression was enhanced in nodules. *M. loti* inoculation or Nod-factor treatment slightly repressed the expression of *CASTOR* and *POLLUX* genes during the first 2 days. This was in contrast to the *NIN* gene, which is also required for nodulation but whose expression was strongly upregulated (Fig. 2d, e).

CASTOR and *POLLUX* encode predicted integral membrane proteins of 853 and 917 amino acids, respectively, with at least four transmembrane (TM) regions (Fig. 3a). Overall the two proteins are very similar to each other with the exception of the amino-terminal regions, which are more divergent (Fig. 3a). However, both proteins carry potential chloroplast transit peptides of 69 and 57 amino acids at their N termini (TargetP scores 0.953 and 0.959), respectively. We confirmed this prediction by using fusions of full-length *CASTOR* or *POLLUX* to green fluorescent protein (GFP). As a positive control, the transit peptide of the *Arabidopsis* plastid protein *recA*¹⁹ (At1g79050) was fused with DsRed2 (*AtrecA*–DsRed2). When this fusion was transiently expressed in onion epidermis or pea root cells together with *CASTOR*–GFP or *POLLUX*–GFP, they localized together, providing strong evidence for the targeting of *CASTOR* and *POLLUX* to plastids (Fig. 4). Given their high level of sequence similarity and their similar localization it is

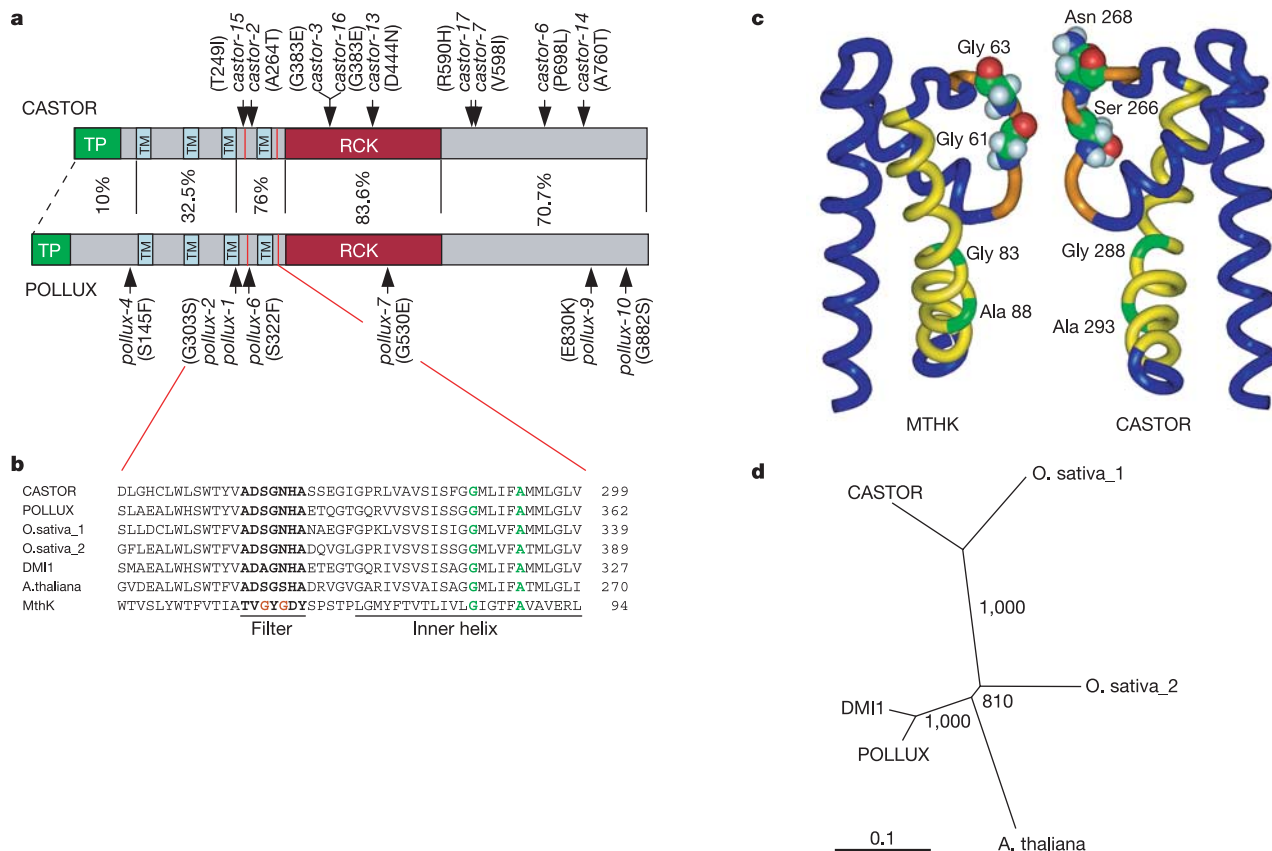


Figure 3 Structure, domains and homologues of *CASTOR* and *POLLUX*. **a**, Protein domains and local sequence identity between *CASTOR* and *POLLUX*. RCK, region with ~35% similarity to a domain implicated in the regulation of conductance, so far only found in potassium channels; TM, transmembrane helix; TP, chloroplast transit peptide. The position and amino-acid changes of non-symbiotic mutations are shown above and below the bars representing the *CASTOR* and *POLLUX* proteins respectively. **b**, Alignment of the filter and inner helical region of the *Methanobacterium thermoautotrophicum* MthK channel (027564) with the homologous region of *CASTOR*, *POLLUX* and plant homologues. This region of MthK forms the core of the pore depicted in **c**. **c**, X-ray crystal structure of the ion pore region of the MthK channel, and a model of the homologous region in *CASTOR*. The filter region is

coloured orange; inner helices are coloured yellow. Residues Gly 61 and Gly 63 in the MthK structure, and the corresponding Ser and Asn residues in *CASTOR*, are shown in space-filling mode to demonstrate the effects of these substitutions on the diameter and electrostatic properties of the pore. Residues Gly 83 (the gating hinge²⁰) and Ala 88 (the narrowest point in the MthK intracellular entryway²⁰) in the inner helix are highlighted in green. **d**, Phylogenetic tree of *CASTOR* and *POLLUX* homologues in angiosperms. The N'-truncated amino-acid sequences of *CASTOR*, *POLLUX* and *DMI1* (*M. truncatula*), and the most similar proteins in *A. thaliana* (At5g49960) and *Oryza sativa* (*O. sativa_1*, deduced from AK068216, and *O. sativa_2*, deduced from AK072312) were aligned with the Clustal W program. Shown is a neighbour-joining tree with 1,000 bootstrap replicates.

surprising that they cannot functionally replace each other, because the mutation of either *CASTOR* or *POLLUX* causes a symbiosis-defective phenotype. One possibility is that *CASTOR* and *POLLUX* form heteromeric complexes.

Database searches based on sequence homology did not result in clear functional predictions for the novel proteins, although similar hypothetical proteins from rice and *Arabidopsis* could be detected (Fig. 3). However, searches based on protein structure using FUGUE²⁰ resulted in significant structural matches to calcium-gated potassium channels such as the *Methanobacterium thermoautotrophicum* MthK, for which a crystal structure has been resolved²¹ (Fig. 3b, c). Comparison with MthK revealed that the TM domains flanking the pore are particularly well conserved in the novel plant proteins (Fig. 3b). However, amino-acid replacements in the filter region suggest different ion selectivity. Residues 61–63 in MthK are Gly-Tyr-Gly, a canonical motif in potassium channels, whereas the corresponding residues in *CASTOR* and *POLLUX* are Ser-Gly-Asn (Fig. 3b). Homology modelling of the filter region of *CASTOR* (Fig. 3c) indicates that these differences result in a more helical backbone, and this would modify both the diameter and electrostatic properties of the pore (Fig. 3c). Despite these observations, it is difficult to predict the ion selectivity of the pore formed by *CASTOR* and *POLLUX*. The proteins also carry a region with homology to the RCK domain, which regulates the conductance of potassium channels in response to changes in cytoplasmic calcium concentration²². As with many other ion channels, MthK acts in a multimeric complex²¹. Given these similarities, it is possible that *CASTOR* and *POLLUX* also act as multimeric ion channels. The significance of these homologies is supported by the observed accumulation of six amino-acid substitutions in the pore and RCK-like domains that lead to a mutant phenotype (Fig. 3a and Supplementary Table 1). However, the channel activity and ion selectivity of *CASTOR* and *POLLUX* have still to be confirmed experimentally.

Our results reveal a new and unexpected aspect of plastid biology. *CASTOR* and *POLLUX* are two closely related members of a novel class of plastid-localized proteins that are absolutely required for

early signal transduction events leading to endosymbioses. How are these proteins implicated in a process that starts with the perception of microbial signals at the plasma membrane and leads to changes in gene expression in the nucleus, among other responses? The immediate-early symbiont-stimulated ion fluxes occur across the plasma membrane⁷, whereas calcium spiking⁴ is typically observed around the nucleus of stimulated root hair cells. Considering their predicted functions and their mutant phenotypes, it seems that *CASTOR* and *POLLUX* might mediate ion fluxes between plastids and cytosol that are a prerequisite for calcium spiking. Although in our experiments, pea plastids seemed to be randomly distributed within root cells, others have observed plastid accumulation around the nucleus²³, so we cannot exclude the plastid as the source of calcium in the spiking response. Determining how the different ion fluxes are orchestrated and how they result in symbiotic changes in root cells will be a challenging research area. Plastids represent the endosymbiotic remnant of a free-living cyanobacterial progenitor. It is fascinating to discover that an 'ancient' endosymbiont helps bacterial and fungal 'newcomers' to infect their partner. □

Methods

Plant material

Mutants *LjSYM71-1*, *LjSYM71-2* (ref. 24), *N5*, *N10*, *LjSYM86-1* (M.K., unpublished data), *LjSYM4-2* (ref. 10), *LjSYM22-1*, *LjSYM23-1*, *LjSYM23-2* (ref. 9) and lines with the prefix 'SL' identified through TILLING²⁵, were isolated from four independent EMS-mutagenesis experiments of *L. japonicus* B-129 Gifu. Non-nodulating mutants *G00472*, *G00716*, *G00862* (derived from B-129) and *M89-27* (from MG-20) were identified from a screen of plants regenerated from cell cultures of *L. japonicus* (Y.U., unpublished data). Mutant *LjSYM4-1* was generated by T-DNA transformation of *L. japonicus* B-129 Gifu, but the gene was not tagged²⁶.

Root hair deformation assay

Root hair deformation was examined as described previously²⁷. In brief, 3-day-old seedlings were inoculated with *M. loti* TONO and grown on a filter paper placed on agar medium (nitrogen-free; B&D²⁸) for 3 days. For Nod factor application, 2-day-old seedlings were grown for 1 day on an agar-covered glass slide with the lower edge immersed in B&D liquid medium, followed by the application of 100 nM Nod factor and incubation for another 24 h. The roots were stained briefly in 0.001% toluidine blue and examined under a binocular microscope. Observation was repeated at least three times, using five to ten seedlings each.

Calcium spiking

Seedlings of *L. japonicus* were germinated, mounted and microinjected with Oregon green-dextran dye essentially as described¹². Fluorescence was imaged with a Nikon TE2000U inverted microscope coupled to a Hamamatsu Photonics digital charge-coupled device camera. The excitation wavelength of 488 nm with an 11-nm bandpass was selected with an Optoscan Monochromator (Cairn Research, Faversham, Kent, UK) and an emission filter of 545 ± 15 nm was used. Images were collected every 5 s with a 200-ms exposure with the use of MetaFluor software, and derivative traces were generated with Microsoft Excel. After microinjection, root hairs were left for at least 20 min before the addition of Nod factor, and only cells showing active cytoplasmic streaming were used for analysis. Nod factors, isolated from a reverse-phase C₁₈ column, were added directly to the incubation chamber to give an estimated final concentration of 10 nM (ref. 29). Representative traces were selected from at least ten independent cells.

Positional cloning

CASTOR was mapped with six independent F₂ mapping populations, resulting from crosses between *L. japonicus* B-129 Gifu mutants *LjSYM71-2*, *N5*, *LjSYM4-2* or *G00472*, *G00716* and MG-20 Miyakojima, and a cross of *LjSYM4-2* and *L. filicaulis*. The sizes of each F₂ population were 1,563 (*LjSYM71-2*), 190 (*N5*), 180 (*LjSYM4-2* × MG-20), 40 (*G00472*), 40 (*G00716*) and 1,514 (*LjSYM4-2* × *L. filicaulis*). In total, 3,527 F₂ individuals were analysed with markers flanking *CASTOR*. *POLLUX* was mapped with three independent F₂ mapping populations. The sizes of each population of F₂ were 531 (*LjSYM23-2* × MG-20), 292 (*LjSYM86-1* × MG-20) and 409 (*LjSYM23-1* × *L. filicaulis*).

Southern blot and expression analysis

A 573-base-pair fragment of *CASTOR* was amplified with the primer 5'-TCAAAGAAGGA TTACGAGGA-3' in combination with 5'-AATTTTACCACCATAAGATGC-3'. Genomic DNA (10 µg) was extracted from leaves and digested with *Xba*I or *Bam*HI. Probe labelling and signal detection were performed with AlkPhos Direct (Amersham).

The expression of the *CASTOR* gene was analysed by quantitative RT-PCR with the primers *polyubiquitin* (5'-ATGCAGATCTTCGTCAAGACCTTGAC-3'/5'-ACCTCCCC TCAGACGAAGGA-3'), *NIN* (ref. 6) (5'-TGGATCAGCTAGCATGGAAT-3'/5'-TCTGCTTCTGCTGTGTGCAC-3'), *CASTOR* (5'-ATGGTGGCCTTGACATAAG-3'/5'-AGTGACGACGTATAACAGCA-3') and *POLLUX* (5'-AGGTATCCTAGGGAAGCAAGC-3'/5'-TTACTCTCTGTTCTCTTCC-3'). Total RNA was extracted with the RNeasy Mini kit (Qiagen) and treated with DNase I (Takara). Real-time RT-PCR was performed in

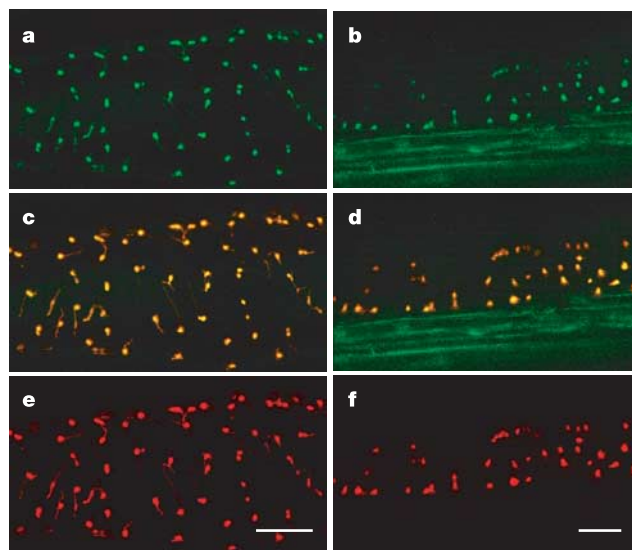


Figure 4 Intracellular localization of *CASTOR* and *POLLUX*. Confocal laser scanning micrographs of onion epidermal cells (**a**, **c**, **e**) and pea roots (**b**, **d**, **f**), transiently expressing *AtrecA*-DsRed2 (**e**, **f**) with *CASTOR*-GFP (**a**) and *POLLUX*-GFP (**b**) fusions, respectively. In the merged images (**c**, **d**), the colocalized signals of green GFP fluorescence (**a**, **b**) and red DsRed2 fluorescence (**e**, **f**) appear yellow. (**a**, **c**, **e**, projection of six optical sections taken at 5-µm intervals along the z axis; **b**, **d**, **f**, projection of eight optical sections taken at 4-µm intervals along the z axis). Green fibrillar cells in **b**, **d** are seen owing to their autofluorescence. Scale bars, 50 µm.

GeneAmp 5700 (Applied Biosystems) with a Quantitect SYBR Green RT-PCR kit (Qiagen). Expression levels were normalized on the basis of the amount of *polyubiquitin* transcripts. All values are means and standard deviations for three triplicates of three biological repeats.

Construction of fluorescence-tagged fusions

cDNA fragments of *CASTOR* and *POLLUX* were amplified by PCR with the primers CAS-F/CAS-R (5'-ACGCGTCGACATGTCCTGGATTCCGAG-3', 5'-CATGCCATGGATTCTTTTCAGTAATTAC-3') and POL-F/POL-R (5'-ACGCGTCGACATGATACCACTACCAGTA-3', 5'-CATGCCATGGAATCGCCTGAAGCAATCAC-3'), respectively. Each fragment was digested with *Sall* and *NcoI* and ligated into the same restriction sites of pUC18-CaMV35S-sGFP (S65T)-nos³⁰. For the construction of *Atreca*-DsRed2 fusion, a fragment encoding the transit peptide of *Atreca*¹⁹ was amplified from *Arabidopsis thaliana* genomic DNA with the use of the primers *Atreca*-*Sall*-F/*Atreca*-LFH-R (5'-ACGCGTCGACATGGATTACACAGCTAGTC-3', 5'-ATCGAATTCAGAACTGATTTTGTG-3'). DsRed2 fragment was amplified from pDsRed2-1 (Clontech) with the use of DsRed2-LFH-F/DsRed2-Nofl-R (5'-CTGAATTCGATCGCGACATGGCCTCTCCGAGAA-3', 5'-ATTTCGGCCGCTACAGGAACAGGTGGTG-3') primers. *Atreca*-LFH-R and DsRed2-LFH-F primers were designed to introduce overlapping nucleotides (underlined) at the 3' and 5' ends of the *Atreca* and DsRed2 fragments, respectively. Using both fragments as templates, joint PCR was performed with *Atreca*-*Sall*-F and DsRed2-Nofl-R primers. The resulting fusion fragment was digested with *Sall* and *Nofl* and cloned into the same restriction site of pUC18-CaMV35S-sGFP (S65T)-nos vector.

Microprojectile bombardment and confocal laser scanning microscopy

Microprojectile bombardment was performed with a Biolistic PDS-1000/He Particle Delivery System (Bio-Rad). Epidermis of *Allium cepa* scaly bulb and roots of *Pisum sativum* were bombarded with a rupture-disc pressure of 1,100 p.s.i. (~7.6 MPa) at a target distance of 6 cm. At 24–40 h after bombardment, they were analysed with a Bio-Rad Radiance2000 confocal laser scanning microscope. Green fluorescence of GFP and red fluorescence of DsRed2 were excited at 488 nm with an argon laser and collected sequentially with a filter set (HQ530/60 and E570LP). Images of both fluorescences were processed and merged with the Lasersharpp2000 program system (Bio-Rad).

Computer analysis

Sequences were analysed by BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) and GENSCAN version 1.0 (<http://genes.mit.edu/GENSCAN.html>). Clustal W (<http://www.ebi.ac.uk/clustalw/>) was used for multiple alignment and evolutionary relationships. The target peptide and transmembrane regions were predicted by TargetP version 1.01 (<http://www.cbs.dtu.dk/services/TargetP/>) and TMHMM version 2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>). For domain and structure analyses, both Pfam (<http://www.sanger.ac.uk/Software/Pfam/>) and FUGUE v.2.0 (<http://www-cryst.bioc.cam.ac.uk/~fugue/>) were applied.

Received 30 September; accepted 29 November 2004; doi:10.1038/nature03237.
Published online 22 December 2004.

- Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* (Academic, London, 1997).
- Oldroyd, G. E. D. Dissecting symbiosis: developments in Nod factor signal transduction. *Ann. Bot.* **87**, 709–718 (2001).
- Kistner, C. & Parniske, M. Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* **7**, 511–518 (2002).
- Ehrhardt, D., Wais, R. & Long, S. Calcium spiking in plant root hairs responding to Rhizobium nodulation signals. *Cell* **85**, 673–681 (1996).
- Truchet, G. *et al.* Sulphated lipooligosaccharide signals from *Rhizobium meliloti* elicit root nodule organogenesis in alfalfa. *Nature* **351**, 670–673 (1991).
- Radutoiu, S. *et al.* Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**, 585–592 (2003).
- Cárdenas, L. *et al.* Ion changes in legume root hairs responding to Nod factors. *Plant Physiol.* **123**, 443–452 (2000).
- Senoo, K. *et al.* Isolation of two different phenotypes of mycorrhizal mutants in the model legume plant *Lotus japonicus* after EMS-treatment. *Plant Cell Physiol.* **41**, 726–732 (2000).
- Szczygłowski, K. *et al.* Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. *Mol. Plant Microbe Interact.* **11**, 684–697 (1998).
- Bonfante, P. *et al.* The *Lotus japonicus* *LjSym4* gene is required for the successful symbiotic infection of root epidermal cells. *Mol. Plant Microbe Interact.* **13**, 1109–1120 (2000).
- Novero, M. *et al.* Dual requirement of the *LjSym4* gene for mycorrhizal development in epidermal and cortical cells of *Lotus japonicus* roots. *New Phytol.* **154**, 741–749 (2002).
- Harris, J. M., Wais, R. & Long, S. R. Rhizobium-induced calcium spiking in *Lotus japonicus*. *Mol. Plant Microbe Interact.* **16**, 335–341 (2003).
- Hayashi, M. *et al.* Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F2 population. *DNA Res.* **8**, 301–310 (2001).
- Nakamura, Y. *et al.* Structural analysis of a *Lotus japonicus* genome. II. Sequence features and mapping of sixty-five TAC clones which cover the 6.5-mb regions of the genome. *DNA Res.* **9**, 63–70 (2002).
- Kawasaki, S. & Murakami, Y. Genome analysis of *Lotus japonicus*. *J. Plant Res.* **113**, 497–506 (2000).
- Kawaguchi, M. *et al.* Providing the basis for genomics in *Lotus japonicus*: the accessions Miyakojima and Gifu are appropriate crossing partners for genetic analyses. *Mol. Gen. Genomics* **266**, 157–166 (2001).
- Stracke, S. *et al.* A plant receptor-like kinase required for both fungal and bacterial symbiosis. *Nature* **417**, 959–962 (2002).
- Ane, J. M. *et al.* *Medicago truncatula* *DMI1* required for bacterial and fungal symbioses in legumes. *Science* **303**, 1364–1367 (2004).
- Köhler, R. H. *et al.* Exchange of protein molecules through connections between higher plant plastids. *Science* **276**, 2039–2042 (1997).
- Shi, J., Blundell, T. L. & Mizuguchi, K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **310**, 243–257 (2001).

- Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
- Jiang, Y. *et al.* Structure of the RCK domain from the *E. coli* K⁺ channel and demonstration of its presence in the human BK channel. *Neuron* **29**, 593–601 (2001).
- Kwok, E. Y. & Hanson, M. R. Plastids and stromules interact with the nucleus and cell membrane in vascular plants. *Plant Cell Rep.* **23**, 188–195 (2004).
- Kawaguchi, M. *et al.* Root, root hair, and symbiotic mutants of the model legume *Lotus japonicus*. *Mol. Plant Microbe Interact.* **15**, 17–26 (2002).
- Perry, J. A. *et al.* A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiol.* **131**, 866–871 (2003).
- Schauser, L. *et al.* Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *Lotus japonicus*. *Mol. Gen. Genet.* **259**, 414–423 (1998).
- Niwa, S. *et al.* Responses of a model legume *Lotus japonicus* to lipochitin oligosaccharide nodulation factors purified from *Mesorhizobium loti* JRL501. *Mol. Plant Microbe Interact.* **14**, 848–856 (2001).
- Broughton, W. J. & Dilworth, M. Y. Control of leghemoglobin synthesis in snake beans. *Biochem. J.* **125**, 1075–1080 (1971).
- Firmin, J. L. *et al.* Resistance to nodulation of cv. Afghanistan peas is overcome by nodX, which mediates an O-acetylation of the *Rhizobium leguminosarum* lipo-oligosaccharide nodulation factor. *Mol. Microbiol.* **10**, 351–360 (1993).
- Isono, K. *et al.* Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains of σ^{70} factors of bacterial RNA polymerases in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **94**, 14948–14953 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Szczygłowski, J. Webb and J. Stougaard for providing mutant seeds; M. Hayashi for help with mapping; T. Kojima and R. Ohtomo for mycorrhiza analysis; Y. Niwa for providing pUC18-CaMV35S-sGFP (S65T)-nos vector; G. Oldroyd and J. Sun for help with Ca-spiking assays; J. Krüger and B. B. H. Wulff for critical reading of the manuscript; J. Soll for providing the pea root transformation protocol before publication; and M. Durrant for help with modelling the CASTOR pore structure. Part of this work was supported by the fund of Promotion of Basic Research Activities for Innovative Biosciences (BRAIN), and Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency. Research at the Sainsbury Laboratory is funded by the Gatsby Charitable Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.K. (kawasa@nias.affrc.go.jp). The sequences have been deposited at the DNA Data Bank of Japan with the following accession numbers: LJTO2K14a (AP006732), LJTO2K14b (AP006733), LJTO2K14c (AP006734), LJT45I15 (AP006736), LJTO2F11 (AP006737), LJT46G19 (AP006735), LJT45B09a (AP006729), LJT45B09b (AP006730) and LJT45B09c (AP006731); genomic sequences (B-129 Gifu) of CASTOR (AB162016), POLLUX (AB162017), and mRNA sequences (B-129 Gifu) of CASTOR (AB162157) and POLLUX (AB162158).

Interaction network containing conserved and essential protein complexes in *Escherichia coli*

Gareth Butland¹, José Manuel Peregrín-Alvarez², Joyce Li¹, Wehong Yang¹, Xiaochun Yang¹, Veronica Canadien³, Andrei Starostine¹, Dawn Richards³, Bryan Beattie³, Nevan Krogan¹, Michael Davey¹, John Parkinson^{2,4,5}, Jack Greenblatt^{1,3,5} & Andrew Emili^{1,5}

¹Banting and Best Department of Medical Research, University of Toronto, 112 College Street, Toronto, Ontario M5G 1L6, Canada

²Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M4K 1X8, Canada

³Affinin Pharmaceuticals, 100 University Avenue, Toronto, Ontario M5J 1V6, Canada

⁴Department of Biochemistry and ⁵Department of Medical Genetics and Microbiology, University of Toronto, Medical Sciences Building, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada

Proteins often function as components of multi-subunit complexes. Despite its long history as a model organism¹, no large-scale analysis of protein complexes in *Escherichia coli* has yet been reported. To this end, we have targeted DNA cassettes into the *E. coli* chromosome to create carboxy-terminal, affinity-tagged alleles of 1,000 open reading frames (~23% of the

genome). A total of 857 proteins, including 198 of the most highly conserved, soluble non-ribosomal proteins essential in at least one bacterial species, were tagged successfully, whereas 648 could be purified to homogeneity and their interacting protein partners identified by mass spectrometry. An interaction network of protein complexes involved in diverse biological processes was uncovered and validated by sequential rounds of tagging and purification. This network includes many new interactions as well as interactions predicted based solely on genomic inference or limited phenotypic data². This study provides insight into the function of previously uncharacterized bacterial proteins and the overall topology of a microbial interaction network, the core components of which are broadly conserved across Prokaryota.

The yeast-based tandem affinity purification (TAP) procedure for isolating protein complexes makes use of site-specific recombination to introduce a dual tagging cassette into chromosomal loci³. *Escherichia coli* does not readily recombine exogenous linear DNA fragments into its chromosome, but expression of the lambda general recombination system (λ -Red) markedly enhances integration^{4,5}. We adapted one such system⁵ to introduce a DNA cassette, bearing a selectable marker and either the TAP or sequential peptide affinity (SPA) tags^{3,6}, into the C termini of open reading frames (ORFs) in the lysogenic *E. coli* strain DY330 (ref. 6), which harbours λ -Red under control of a temperature-sensitive repressor⁵ (Fig. 1a). The tagged bait proteins, expressed at endogenous levels, were purified $\sim 10^6$ -fold to homogeneity from log-phase cultures using two rounds of affinity chromatography⁶. To minimize nucleic-acid-mediated interactions, extracts were pre-treated with nuclease. Polypeptide components of isolated complexes were then identified using two forms of mass spectrometry: peptide mass fingerprinting was performed on all silver-stained polypeptide bands visible by SDS–polyacrylamide gel electrophoresis (PAGE) not seen in parallel control purifications, whereas gel-free shotgun sequencing was used to identify small and lower-abundance proteins. To minimize the false discovery rate (false positives), protein–protein interactions were deemed authentic provided a reciprocal interaction could be confirmed or if the data were reproducible. Figure 1b shows a schematic overview of the methodology.

The effectiveness of the procedure was confirmed in pilot purifications of DNA-dependent RNA polymerase (RNAP). Tagged core subunit β (RpoB) co-purified specifically with essential elongation factors (NusA and NusG), specialized sigma factors involved in promoter recognition (σ^{32} (RpoH), σ^{38} (RpoS), σ^{54} (RpoN), σ^{70} (RpoD)), and with accessory factors ω (RpoZ), HepA (RapA) and YacL (15 kDa acidic protein of unknown function not previously known to interact with RNAP) (Fig. 2a; see also Supplementary Table 1). Similarly, NusG co-purified with YacL, HepA, core enzyme and termination factor Rho, whereas ω bound σ^{70} , NusA and b1731, another small protein of unknown function. In reciprocal experiments, tagged b1731 co-purified with β , β' (RpoC), α (RpoA), σ^{70} and ω , but not with Nus factors, HepA or YacL (Fig. 2b), implying an exclusive association with initiating holoenzyme. In contrast, tagged YacL bound ω , NusG and HepA together with core enzyme (Fig. 2c; see also Supplementary Table 1), suggesting a role in elongation. On the basis of their specific and reproducible association with RNAP, we suggest that b1731 and YacL be renamed Rap (RNAP-associated protein) B and C, respectively. YacL and b1731 are not required for viability, and homologues are restricted to γ -Proteobacteria (Supplementary Fig. 1a), suggesting a specialized function. Notably, tagged σ^{54} and σ^{38} bound to the sugar transporter ManX, whereas σ^{32} co-purified with quinone oxidoreductase Qor, indicating that protein–protein interactions may regulate alternative sigma factor activities.

Purification of DNA-dependent DNA polymerase indicated that low-abundance complexes are amenable to analysis. A nine-subunit holo-complex, DNA polymerase III*, which contains a core complex (α (DnaE), ϵ (DnaQ) and θ (HolE)) requiring only processivity

factor β (DnaN) for full replicative activity⁷, was readily isolated (Supplementary Table 1). Purification of tagged α , ϵ and θ yielded core complex as well as the clamp loader (γ and τ (dnaX), δ (HolA), δ' (HolB), χ (HolC) and ψ (HolD)), which recruits β onto DNA (Fig. 2d; θ detected by gel-free mass spectrometry). In reciprocal experiments (Supplementary Table 1), tagged χ and ψ co-purified with core enzyme and the clamp loader. χ also bound PriA (primosomal DNA helicase), b1808 (putative ATP-dependent helicase) and TopA (DNA topoisomerase I), along with three conserved replication factors that act coordinately *in vitro*⁸: Ssb, single-strand DNA-binding protein previously reported to bind χ ⁹; RecQ, a replicative DNA helicase; and TopB, DNA topoisomerase III (Fig. 2d). In turn, Ssb co-purified with TopB, exonucleases RecJ and SbcB, and helicases PriA, RecG and RecQ (Supplementary Table 1), consistent with multiple roles in chromosome dynamics.

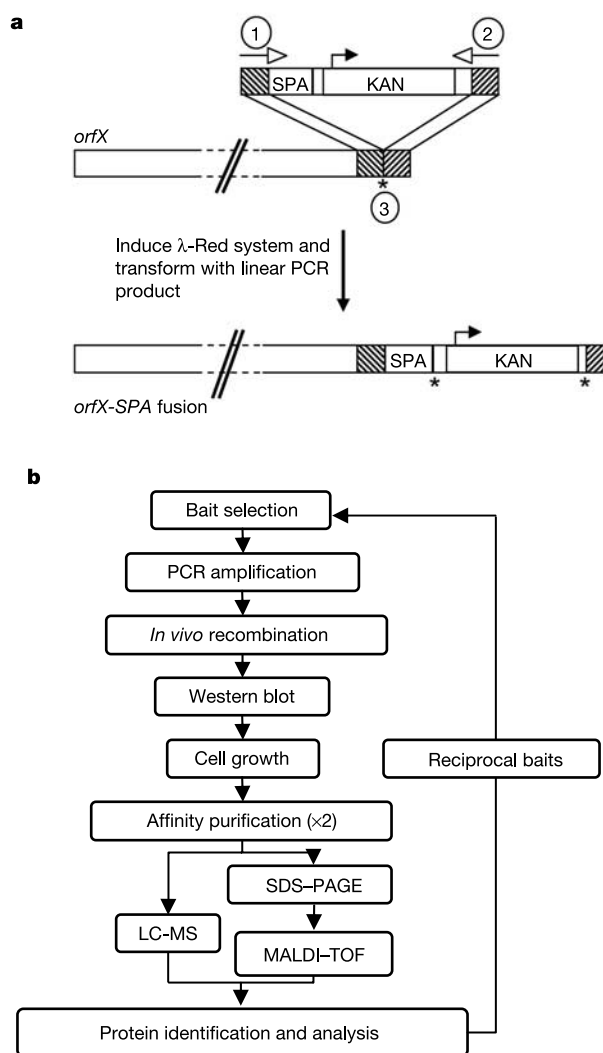


Figure 1 Systematic identification and validation of protein complexes in *E. coli*. **a**, Gene-specific affinity-tagging cassettes produced by polymerase chain reaction (PCR)⁶ using primers (1, 2) homologous to a target translational termination codon (3) were integrated into the *E. coli* chromosome using the λ -Red recombination system⁵. KAN, kanamycin-resistance cassette. Asterisks indicate stop codons. **b**, Flow chart of steps in purifying and validating protein complexes. Interactions were confirmed either by reciprocal tagging and purification, or by repeat analysis. LC-MS, gel-free liquid chromatography–tandem mass spectrometry; MALDI-TOF, gel-based peptide mass fingerprinting using matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry.

TopB co-purified with Ssb and RecQ, suggesting that these factors serve a similar role in coordinating DNA replication in prokaryotes as suggested for their eukaryotic homologues¹⁰.

We expanded our analysis to cover about one-quarter of the *E. coli* genome, targeting 1,000 ORFs (Supplementary Fig. 2 and Supplementary Table 2), including 248 uncharacterized 'y' genes, 168 putative 'b' genes and 209 of the most broadly conserved genes encoding soluble (non-membrane) non-ribosomal proteins essential in *E. coli* or another bacterium (see Methods). We successfully tagged 857 proteins (86%; confirmed by western blot analysis), including 198 essential and conserved (essential-conserved) proteins, and were able to purify 648 (65%; detected by mass spectrometry), which compares favourably to analogous studies of protein complexes in yeast³. A total of 118 of these proteins had no detectable partners, whereas 5,254 putative protein-protein interactions were detected for the other 530 baits (Supplementary Table 1). To eliminate false-positives, we reciprocally tagged and purified a large subset of candidate partners. A validation rate of ~53% was achieved (see Supplementary Information), which compares favourably to studies in yeast¹¹, confirming the stringency of the methodology. A total of 716 non-redundant binary interactions, involving 83 essential (excluding the ribosome) and 152 non-essential proteins, have been validated so far (highlighted in Supplementary Table 1). The entire validated data set is shown graphically in Fig. 3a.

Eighty-five per cent of the validated interactions are new, as they are not described in the Database of Interacting Proteins (DIP)¹², Biomolecular Interaction Network Database (BIND)¹³, STRING¹⁴, or Prolinks databases¹⁵, whereas only ten orthologous interactions (interologs) were reported in a two-hybrid interaction screen in *Helicobacter Pylori*¹⁶ (Supplementary Table 3). The significance of these novel interactions is reinforced by functional annotation (Supplementary Table 2). For instance, acyl carrier protein (ACP), a key carrier of growing fatty acid chains, bound specifically and reproducibly to enzymes linked to biogenesis of fatty acids, phospholipids and lipid A (essential outer-membrane constituent) (Fig. 2e; see also Supplementary Table 1), including two 3-ketoacyl-ACP synthases (FabB, FabF), 3-ketoacyl-ACP reductase (FabG), 3-hydroxyacyl-ACP dehydrase (FabZ), LpxD (essential protein

required for lipid A biogenesis), YbgC (*tol-pal* cluster hydrolase of short-chain acyl-CoA thioesters), AcpS (involved in transfer of 4'-phosphopantethein to ACP), Aas and PlsB (membrane proteins involved in phospholipid acylation), and YiiD (putative acetyltransferase). ACP also co-purified with GlmU (an essential bi-functional enzyme that converts glucosamine-1-phosphate to UDP-GlcNAc (lipid A precursor)), AidB (isovaleryl-CoA dehydrogenase), SecA (pre-protein translocase), as well as MukB and SpoT, as previously reported¹⁷. We did not detect IscS, which was predicted to interact with ACP in an overexpression study¹⁷, possibly because transient or low-affinity interactions are missed by our method.

Many other informative complexes were detected (see Supplementary Information). These included notable interactions mediated by the cysteine desulfurase IscS (IscS-FdhD and IscS-YhhP), between two uncharacterized essential proteins (YgjD-YeaZ), and by a sizeable group of uncharacterized proteins with factors involved in ribosome function, RNA processing, and/or RNA binding.

Graph network analysis of the validated data set provided evidence of 'scale-free' behaviour¹⁸. Most of the proteins had few interacting partners, whereas a subset of 'hubs' formed a far greater number of connections (Fig. 3b). Comparable connectivity was observed for the essential-conserved proteins alone (Supplementary Fig. 3). Scale-free networks are predicted to be robust against random node removal but vulnerable to hub removal, a property that might be expected to be preserved across evolution¹⁸. Indeed, removal of the 20 most highly connected nodes (>15 interactions; Supplementary Table 4) markedly reduced the network connectivity (see Supplementary Figs 4a and b). Notably, these same hubs were all highly conserved (detected in ≥125 genomes; Supplementary Fig. 4c). Moreover, protein connectivity was proportional (positively correlated) to the number of genomes a homologue was detected in (Supplementary Fig. 4d). Although a previous analysis of bacterial two-hybrid data¹⁹ failed to detect such a dependency, possibly due to a high false-positive rate, our results are in agreement with a more in-depth analysis of the relationship of protein evolutionary rates to the number of interactions in eukaryotes²⁰.

To investigate further the conserved nature of the bacterial network, we analysed co-occurrence of BLAST homologues of

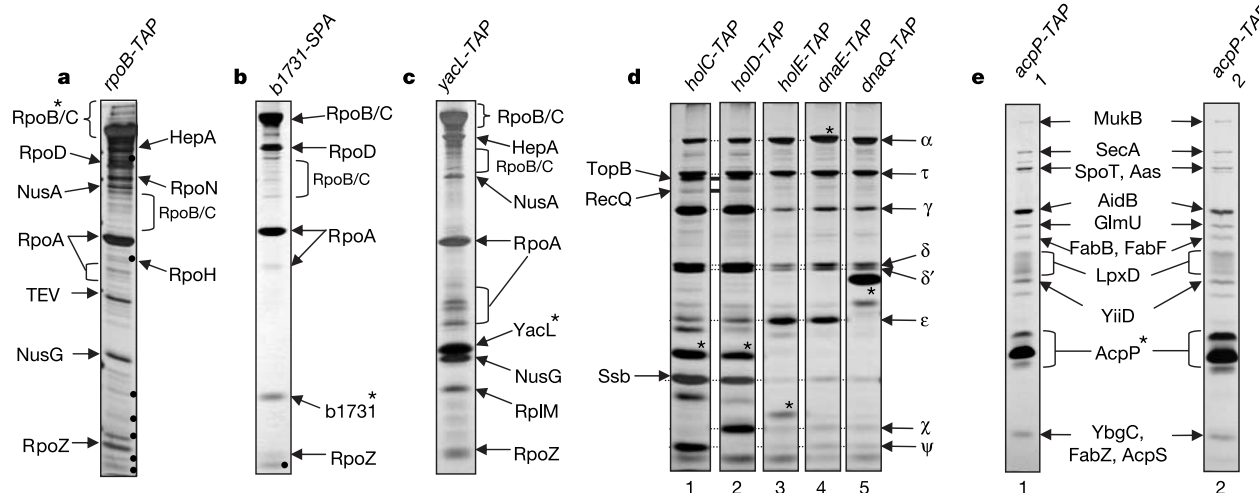


Figure 2 Analysis of affinity-purified protein complexes. SDS-PAGE silver-stain analysis of the components of affinity-purified complexes from *E. coli*. **a–c**, Purification of TAP-tagged *E. coli* RNAP subunit β (**a**) and two associated proteins: SPA-tagged b1731 (**b**) and TAP-tagged YacL (**c**). **d**, Purification of TAP-tagged HoID (ψ) and HoIC (χ) subunits of the processivity clamp loader (lanes 1 and 2), and DNA polymerase III core subunits

HoIE (θ), DnaE (α) and DnaQ (ε) (lanes 3–5). **e**, Independent purifications of TAP-tagged AcpP (lanes 1 and 2). Only validated interacting proteins are labelled. Black circles indicate bands that failed to yield spectra or validated results; asterisks indicate tagged bait; and brackets indicate degradation products.

each pair of interacting proteins across all three domains of life (Archaea, Prokaryota and Eukaryota)²¹ (see Methods). As seen in Supplementary Fig. 5, the interacting proteins were more likely to be co-conserved than control, randomly selected protein pairs,

indicating that the interactions are similarly conserved. Notably, the most highly conserved proteins were highly connected, forming a single interconnected component (Fig. 3c, d). This core set of 154 interactions involving 71 proteins (including the ribosome;

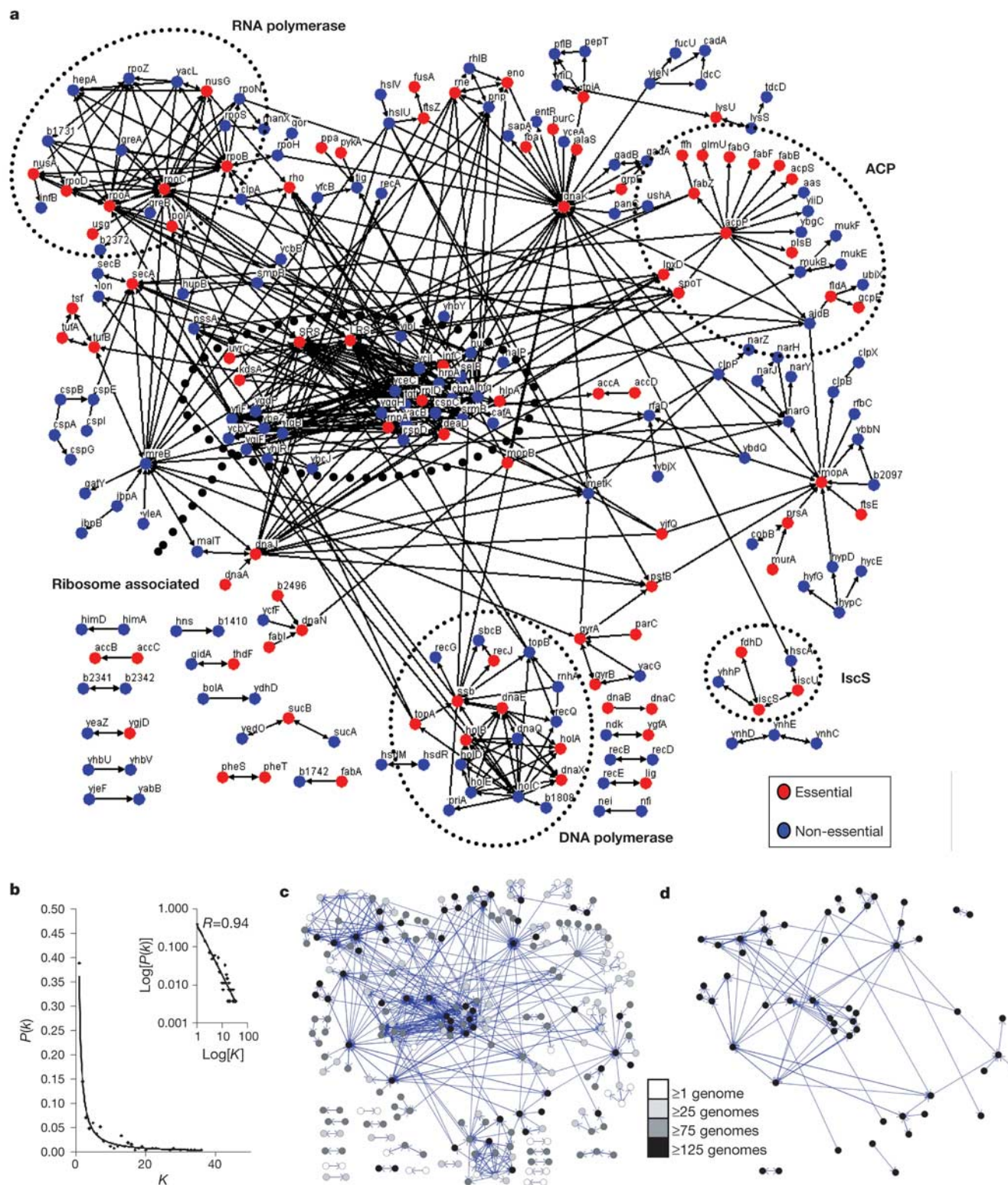


Figure 3 Network properties of bacterial protein–protein interactions. **a**, Network of validated protein complexes. Interactions are represented as directional edges extending from the tagged protein. Baits without partners are removed for clarity. Red nodes, essential proteins; blue nodes, non-essential proteins; black ovals, complexes discussed in text. **b**, Connectivity distribution of validated interactions (K) per protein plotted as a

function of frequency, $P(k)$. Inset: log-plot power law distribution, $P(k) \approx k^{-\gamma}$, where γ is the degree exponent. R , Pearson's correlation coefficient (see Methods). **c**, Network sequence conservation. Node shading (white-to-black) is scaled according to the number of genomes that pairs of interacting proteins co-occur in. **d**, Network of highly conserved proteins co-occurring in ≥ 125 genomes (homologue raw BLAST bit score ≥ 50).

Supplementary Table 5) potentially fulfils critical roles across all bacteria. Similar results were obtained using clusters of orthologous groups of proteins²² (COGs; Supplementary Fig. 6a, b).

Co-occurrence of homologues across different genomes (phylogenetic profiles) has previously been used to explore functional links between genes^{23,24}. Different approaches for constructing these profiles include the use of orthology assignments (for example, COGs) and sequence homology²⁵. Recently, a study²⁶ introduced a mapping approach based on sequence homology methods to assess the degree of conservation of interologs between species. We adopted a similar approach to examine patterns of interolog conservation within the complexes detected in this study. Intriguingly, AcpP and several of its interacting partners displayed significant divergence in Bacilli, Actinobacteridae, *Mycoplasma* spp. as well as Archaea and eukaryotes (Fig. 4a). The lack of obvious homologues of AcpP in actinobacteria is consistent with the highly diverged nature of the predicted orthologues from this phylum in the COGs classification scheme. Evolutionary divergence was also evident with the DNA and RNA polymerase complexes. Although core RNAP subunits and cofactors *nusA*, *rpoA*, *rpoB*, *rpoC*, *rpoD*, *rpoH* and *rpoS* are found in virtually all eubacteria, other subunits

(such as *hepA*, *rpoZ*, *b1731* and *yacL*) are restricted to γ -Proteobacteria (Supplementary Fig. 1a). Likewise, the DNA polymerase clamp loader subcomplex (*hola-D*) is similarly restricted (Supplementary Fig. 1b). These data suggest that sequence divergence may lead to functional diversification and the formation of novel modules. By clustering interacting proteins based on their phylogenetic profiles, it may be possible to identify new modules. For instance, the interacting PflB–PepT gene products cluster together (see Supplementary Fig. 2).

Statistical methods such as mutual information, Pearson correlation coefficient, Hamming distance (D) and the chance co-occurrence probability distribution (P) have been developed to predict functional relationships among genes based on phylogenetic profiles²³. We applied the latter two metrics to quantify the extent of correlation between different phylogenetic profiles. Using relatively modest cutoff values ($D < 25$, $P < 10^{-11}$), a small but significant subset ($\sim 14\%$) of the interacting proteins showed closely correlated phylogenetic profiles relative to a control set of randomly selected pairs of bacterial proteins (Fig. 4b). Chromosomal proximity has also been used to infer functional linkages between evolutionarily conserved proteins²⁷. Investigations into the distribution of relative

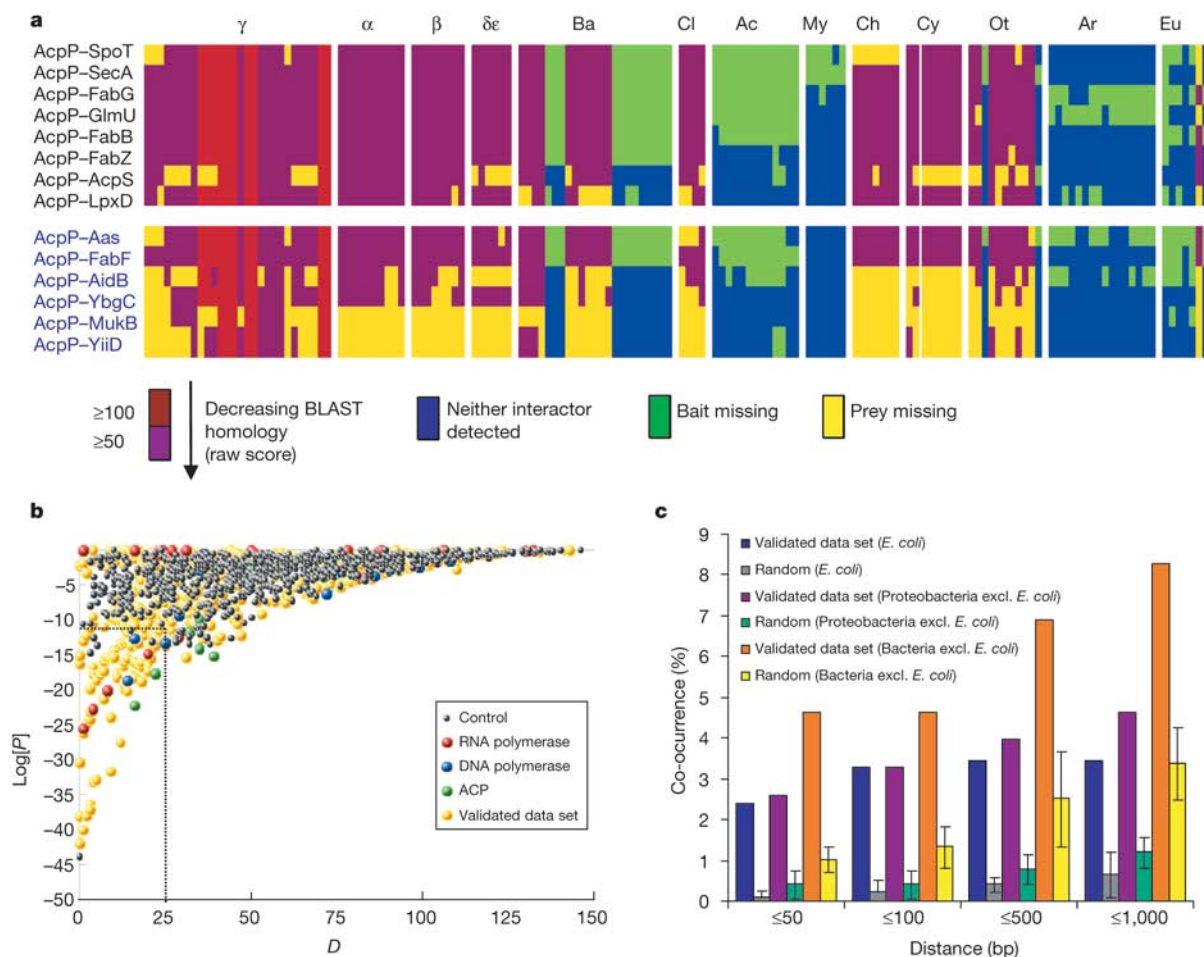


Figure 4 Bioinformatic analyses of interacting protein modules. **a**, Phylogenetic profile of ACP interactions (bait–partner) within 148 genomes. Coloured boxes indicate degree of BLAST homology. Black font, bait–partner both essential-conserved proteins; blue font, only bait is an essential-conserved protein. Phylogenies from the NCBI taxonomy database (see Supplementary Information). γ , α , β and $\delta\epsilon$ indicate respective proteobacteria. Ba, Bacilli; Cl, Clostridia; Ac, Actinobacteridae; My, *Mycoplasma*; Ch, Chlamydiae; Cy, Cyanobacteria; Ot, unclassified bacteria; Ar, Archaea; Eu, Eukaryota. **b**, Scatter log-plot of Hamming distance (D) versus chance co-occurrence probability distribution (P) for

interactions associated with RNAP, DNA polymerase, ACP, the complete validated data set or random protein pairs (control). The dotted box indicates interactions ($\sim 14\%$) with significantly correlated profiles. **c**, Bar graph of relative genomic distances (base pairs) between genes encoding interacting proteins. Orange/yellow, orthologues in bacteria (excluding *E. coli*; see Supplementary Fig. 1); purple/green, orthologues in proteobacteria (excluding *E. coli*); blue/grey, *E. coli*. Error bars indicate standard deviation for 20 replicate controls.

chromosomal distances between loci encoding pairs of interacting proteins revealed that only a modest proportion (3.4%) of the interacting proteins were encoded by genes located within 500 base pairs (bp) of each other within the *E. coli* genome (Fig. 4c). A slightly larger fraction (6.9%) of putative orthologues was similarly separated by less than 500 bp in at least one other bacterial genome (Fig. 4c; see also Supplementary Table 6). Importantly, of the 42 protein pairs satisfying this criterion, 18 are encoded within the same operon in *E. coli* (see Supplementary Information), including 5 of the 15 interactions involving AcpP, further validating these data. These results indicate that only a fraction of our experimentally detected bacterial protein interactions could be readily predicted by genome-context methods (see above and Supplementary Table 3).

In summary, a reliable network of functionally diverse protein complexes was elucidated in *E. coli* using an experimental approach that can be readily adapted to other prokaryotes⁶. These data offer an insight into the function of uncharacterized proteins and outline the topological organization of a bacterial interactome. Because only about 30 bacterial proteins are currently targeted by prescription drugs²⁸, knowledge of physical interactions mediated by conserved, essential bacterial proteins should facilitate the design of broad-range antimicrobials. These data should also prove valuable for calibrating computational approaches designed to predict functional associations between proteins. Moreover, the tagged strain collection should facilitate biochemical studies using traditional or microarray-based assays. □

Methods

Construction of TAP/SPA-tagged *E. coli* strains

Escherichia coli strains bearing either TAP- or SPA-tagged alleles were constructed by targeted homologous recombination of DNA cassettes into the *E. coli* strain DY330 as previously reported⁶. Primer sequences are available upon request. The SPA tag was selected for use after initial trials due to the reduced protein degradation observed when compared with TAP-purified protein complexes.

Large-scale SPA/TAP purification of *E. coli* protein complexes

TAP or SPA purification was performed using 2–4 l log-phase cultures as previously described⁶ except that Benzonase nuclease (Novagen; 3 U) was incubated with the cleared cell extract on ice for 30 min before purification. Purified complexes were split into aliquots for SDS-PAGE and liquid chromatography-tandem mass spectrometry analysis as detailed in the text.

Identification of proteins by mass spectrometry

Complex subunits were separated by SDS-PAGE on 12% acrylamide gels with a Whatman V16 vertical gel apparatus run at low current (9 mA) for 16 h. Gels were silver-stained using a standard protocol, except that formaldehyde crosslinking was not performed (details available upon request). Protein bands were excised and analysed as described previously²⁹. Gel-free protein sequencing was performed by microcapillary-scale liquid chromatography-electrospray-ion trap tandem mass spectrometry as described²⁹. Spectra were searched against an in-house database of predicted *E. coli* protein-coding sequences.

Selection of target gene products

ORFs were selected for study to obtain broad biological coverage, including highly conserved essential and non-essential proteins, proteins with putative functional assignments, and hypothetical uncharacterized ORFs. Proteins known or predicted to contain trans-membrane helices were avoided owing to technical difficulties associated with purifying membrane proteins. The essential-conserved set of genes was selected using a basic rule set (see Supplementary Information).

Connectivity distribution

To assess the correlation between connectivity (k) and frequency ($P(k)$), and between connectivity and the number of genomes a homologue was detected in, we calculated the Pearson's correlation coefficient R .

Phylogenetic analysis of protein complexes

For each *E. coli* sequence, a TBLASTN³⁰ search was performed against each of the different organism genome data sets. In addition, to avoid complications caused by intronic regions, a protein data set was obtained for each eukaryotic organism considered here and a BLASTP³⁰ performed. The raw score for the highest sequence match to each data set was extracted and stored in a local database. Phylogenetic interaction profiles were visualized using a java applet developed in house.

Phylogenetic distribution of proteins and interactions

The phylogenetic distribution of the proteins was obtained as previously described²¹ using

as reference the non-redundant protein sequence database SWALL (SwissProt plus TrEMBL; see Supplementary Methods). $P(h)$ represents the frequency of proteins with a homologue in the corresponding taxonomic groups. We used BLAST³⁰ with a threshold raw score of 50 and default parameters. We considered a protein interaction to be conserved if both interacting proteins have detectable homologues in any of the 148 complete genomes analysed (see legend to Supplementary Fig. 1). To analyse statistical significance, 30 control set samples of equal size were taken from the *E. coli* genome as previously described²¹. We used a two-tailed t -test, at a 95% confidence level.

Phylogenetic profiles

Each gene is represented by a vector representing the pattern of co-occurrence across 148 genomes (with a value of 1 assigned when a homologue is present, and 0 when one is not). The extent of correlation between phylogenetic profiles of pairs of interacting proteins was assessed by computing Hamming distance (D) and the chance co-occurrence probability distribution (P)³⁴. If N is the total number of genomes over which we construct a phylogenetic profile for R genes, X and Y are the number of genomes in which homologues of two genes occur, and z is the number of genomes in which the genes co-occur, then $P = w_z \bar{w}_z / W$, where w_z is the number of ways to distribute z co-occurrences over N genomes, $\bar{w}_z$ is the number of ways to distribute the rest of the $X - z$ and $Y - z$ genes over the rest of the $N - z$ lineages, and W is the number of ways to distribute X and Y over N genomes without restriction, and $D = x + y - 2z$.

Genomic distance

Relative distances between genes encoding each pair of interacting proteins were calculated using the chromosome location coordinates of genes in COGs²².

Received 6 October; accepted 3 December 2004; doi:10.1038/nature03239.

- Neidhardt, F. (ed.) *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ASM Press, Washington DC, 1996).
- Serres, M. H. et al. A functional update of the *Escherichia coli* K-12 genome. *Genome Biol.* **2**, research0035.1–0035.7 (2001).
- Gavin, A. C. et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
- Yu, D. et al. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **97**, 5978–5983 (2000).
- Zeghouf, M. et al. Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes. *J. Proteome Res.* **3**, 463–468 (2004).
- Stukenberg, P. T. & O'Donnell, M. Assembly of a chromosomal replication machine: two DNA polymerases, a clamp loader, and sliding clamps in one holoenzyme particle. V. Four different polymerase-clamp complexes on DNA. *J. Biol. Chem.* **270**, 13384–13391 (1995).
- Harmon, F. G., Brockman, J. P. & Kowalczykowski, S. C. RecQ helicase stimulates both DNA catenation and changes in DNA topology by topoisomerase III. *J. Biol. Chem.* **278**, 42668–42678 (2003).
- Witte, G., Urbanek, C. & Curth, U. DNA polymerase III chi subunit ties single-stranded DNA binding protein to the bacterial replication machinery. *Nucleic Acids Res.* **31**, 4434–4440 (2003).
- Nakayama, H. RecQ family helicases: roles as tumor suppressor proteins. *Oncogene* **21**, 9008–9021 (2002).
- von Mering, C. et al. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**, 399–403 (2002).
- Salwinski, L. et al. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **32** (Database issue), D449–D451 (2004).
- Bader, G. D., Betel, D. & Hogue, C. W. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31**, 248–250 (2003).
- von Mering, C. et al. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* **31**, 258–261 (2003).
- Bowers, P. et al. Prolinks: a database of protein functional linkages derived from coevolution. *Genome Biol.* **5**, R35 (2004).
- Rain, J. C. et al. The protein–protein interaction map of *Helicobacter pylori*. *Nature* **409**, 211–215 (2001).
- Gully, D., Moir, D., Loiseau, L. & Bouveret, E. New partners of acyl carrier protein detected in *Escherichia coli* by tandem affinity purification. *FEBS Lett.* **548**, 90–96 (2003).
- Wuchty, S., Oltvai, Z. N. & Barabási, A. L. Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nature Genet.* **35**, 176–179 (2003).
- Jordan, I. K., Wolf, Y. I. & Koonin, E. V. No simple dependence between protein evolution rate and the number of protein–protein interactions: only the most prolific interactors tend to evolve slowly. *BMC Evol. Biol.* **3**, 1 (2003).
- Fraser, H. B., Wall, D. P. & Hirsh, A. E. A simple dependence between protein evolution rate and the number of protein–protein interactions. *BMC Evol. Biol.* **3**, 11 (2003).
- Peregrin-Alvarez, J. M., Tsoka, S. & Ouzounis, C. A. The phylogenetic extent of metabolic enzymes and pathways. *Genome Res.* **13**, 422–427 (2003).
- Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278**, 631–637 (1997).
- Wu, J., Kasif, S. & DeLisi, C. Identification of functional links between genes using phylogenetic profiles. *Bioinformatics* **19**, 1524–1530 (2003).
- Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl Acad. Sci. USA* **96**, 4285–4288 (1999).
- Date, S. V. & Marcotte, E. M. Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages. *Nature Biotechnol.* **21**, 1055–1062 (2003).
- Yu, H. et al. Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs. *Genome Res.* **14**, 1107–1118 (2004).
- Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. The use of gene clusters to infer

- functional coupling. *Proc. Natl Acad. Sci. USA* **96**, 2896–2901 (1999).
28. Haselbeck, R. *et al.* Comprehensive essential gene identification as a platform for novel anti-infective drug discovery. *Curr. Pharm. Des.* **8**, 1155–1172 (2002).
29. Krogan, N. J. *et al.* High-definition macromolecular composition of yeast RNA-processing complexes. *Mol. Cell* **13**, 225–239 (2004).
30. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors wish to thank C. J. Ingles and M. Shales for comments on the manuscript. This work was supported by funds from the Ontario Research and Development Challenge Fund and Genome Canada to A.E. and J.G. G.B. was a recipient of a Charles H. Best Post-Doctoral Fellowship. J.M.P.-A. acknowledges support from the Hospital for Sick Children (Toronto, Ontario, Canada) Research Training Centre. Computer analyses were undertaken at the Centre for Computational Biology, Hospital for Sick Children.

Authors' contributions Informatics studies were performed and analysed by J.M.P.-A. and J.P. Experimental design and data analysis were coordinated by G.B. Tagging and purification experiments were performed by W.Y., X.Y., J.L. and G.B. V.C., A.S., D.R., B.B., N.J.K. and M.D. performed and assisted with mass spectrometry analysis. The manuscript was jointly drafted by G.B., A.E., J.G., J.M.P.-A. and J.P. The project was conceived and designed by J.G. and was directed by A.E.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.E. (andrew.emili@utoronto.ca) or J.G. (jack.greenblatte@utoronto.ca).

Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5' - and 3' -halves

Lennart Randau^{1,2}, Richard Münch², Michael J. Hohn^{1,3}, Dieter Jahn² & Dieter Söll^{1,4}

¹Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, USA

²Institut für Mikrobiologie, Technical University Braunschweig, P.O. Box 3329, D-38023 Braunschweig, Germany

³Lehrstuhl für Mikrobiologie und Archaeenzentrum, Universität Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany

⁴Department of Chemistry, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, USA

Analysis of the genome sequence of the small hyperthermophilic archaeal parasite *Nanoarchaeum equitans*^{1,2} has not revealed genes encoding the glutamate, histidine, tryptophan and initiator methionine transfer RNA species. Here we develop a computational approach to genome analysis that searches for widely separated genes encoding tRNA halves that, on the basis of structural prediction, could form intact tRNA molecules. A search of the *N. equitans* genome reveals nine genes that encode tRNA halves; together they account for the missing tRNA genes. The tRNA sequences are split after the anticodon-adjacent position 37, the normal location of tRNA introns. The terminal sequences can be accommodated in an intervening sequence that includes a 12–14-nucleotide GC-rich RNA duplex between the end of the 5' tRNA half and the beginning of the 3' tRNA half. Reverse transcriptase polymerase chain reaction and aminoacylation experiments of *N. equitans* tRNA demonstrated maturation to full-size tRNA and acceptor activity of the tRNA^{His} and tRNA^{Glu} species predicted *in silico*. As the joining mechanism possibly involves tRNA *trans*-splicing, the presence of an intron might have been required for early tRNA synthesis.

The origin of the tRNA molecule is the subject of continuing

discussions and has led to different models postulating that tRNA evolved by duplication or ligation of an RNA hairpin^{3,4}. To examine these models further, the investigation of ancient tRNA genes was central. An interesting organism for this task was *Nanoarchaeum equitans*, currently the only characterized member of the kingdom Nanoarchaeota, which roots early in the archaeal lineage, before the emergence of Euryarchaeota and Crenarchaeota⁵. A significant fraction of the small number of *N. equitans* open reading frames consists of 'split genes' that are encoded as fused versions in other archaeal genomes. Our attention was caught by the 'absence' of four tRNA genes encoding the glutamate, histidine, tryptophan and initiator methionine acceptors⁶.

We therefore developed a computational approach to search for tRNA signature sequences in the *N. equitans* genome. Our program, trained by an alignment of 4,000 tRNA gene sequences (taken from ref. 6), identifies sequences comprising the highly conserved T-loop region and defines the adjacent 3'-acceptor stem sequence. The reverse complementary sequence (defining the 5'-acceptor stem sequence) plus a D-stem position weight matrix identifies the corresponding 5' half. The length of the position weight matrices can be adjusted and mismatches in the acceptor stem can be included. Finally, putative tRNA-halves are ligated *in silico* and analysed by COVE⁷. In addition to identifying the set of tRNAs predicted by the tRNAscan-SE program⁸, our algorithm found nine tRNA halves spread throughout the chromosome. Surprisingly, these tRNA halves could be joined *in silico* to form the missing tRNA^{His}, tRNA^{Met}, tRNA^{Trp} and two tRNA^{Glu} species (Fig. 1). Further analysis of the tRNA half genes revealed several striking features. First, the location of the sequence separation that generated all nine tRNA half genes is after position 37, one nucleotide downstream of the anticodon and the common location of tRNA introns⁹. Second, a consensus sequence matching the highly conserved archaeal Box A promoter element¹⁰ was found upstream of all 5' tRNA halves. Third, this same consensus sequence (5'-TTTT/ATAAA-3') was located 17–25 base pairs (bp) further upstream of the 3' tRNA halves, resulting in a transcript with a 12–14-bp-long GC-rich leader sequence. Last, it is remarkable that this leading sequence is in all cases the exact reverse complement to a sequence following the corresponding 5' tRNA half.

The existence of three tRNA^{Glu} half genes was most exciting. Two 5' tRNA halves were identified that differed solely by one anticodon base (isoacceptors with UUC and CUC anticodon), whereas only one 3' tRNA^{Glu} half gene was found. Both 5' tRNA^{Glu} half genes were followed by the identical 14-bp sequence that was the exact reverse complement of the single 3' tRNA^{Glu} half upstream sequence. All identified split tRNA genes contained the consensus bases of all archaeal elongator tRNAs⁶, namely U8, A14, G15, G18, G19, C32, U33 and the T-loop GTTCA/GAATC (53–61), with the exception of the putative tRNA^{Trp} harbouring an unusual GG sequence preceding the anticodon. The identified tRNA^{Met} displays the consensus sequences of archaeal initiator tRNAs such as the anticodon stem/loop nucleotides (nt) 29–41 (GGGCU-CAUAACCC) and the R11:Y24 base pair (G11:C24), which is the reverse of the Y11:R24 base pair found in elongator tRNAs including the annotated *N. equitans* tRNA^{Met}. Therefore we define the split tRNA^{Met} as the missing initiator tRNA. The sequences also reveal characteristic nucleotides in the respective tRNA species needed for recognition by the cognate aminoacyl-tRNA synthetase. For example, the tRNA^{His} half genes encode the unique G-1:C73 base pair required for aminoacylation of tRNA^{His} by histidyl-tRNA synthetase¹¹, and the tRNA^{Glu} isoacceptors contain the characteristic D-loop nucleotides 20a and 20b and the deletion of base 47 essential for making the 'augmented D-helix'¹².

We performed reverse transcriptase polymerase chain reaction (RT-PCR) analysis of *N. equitans* total tRNA to verify the computationally predicted sequence of the newly discovered joined tRNAs. Our sequencing results confirmed the sequences for tRNA^{Glu}

(UUC), tRNA^{Glu} (CUC) and tRNA^{Met} (Fig. 2a). Despite extensive efforts we could not amplify the full-length tRNA^{Trp} and tRNA^{His} (even though its existence was shown by aminoacylation; see below); this might have been due to the extreme thermostability of the GC-rich *N. equitans* tRNAs containing modified nucleosides. Nevertheless, we confirmed the presence of six tRNA half transcripts by RT-PCR and sequence analysis (Fig. 2d). The primary transcripts of these tRNA half genes include the intervening complementary sequences at the position of separation. In addition, RT-PCR of anchor-ligated tRNA (Fig. 2c) revealed that the primary transcript of the 5' tRNA^{His} half terminates at the AT-rich region following the complementary downstream sequence found in all tRNA half genes.

For a tRNA to participate in protein biosynthesis it must carry a 3'-terminal CCA sequence to which the amino acid will be esterified. In *N. equitans* and most Archaea, this CCA sequence is not encoded in the tRNA genes (including the split tRNA genes) but is added post-transcriptionally by the ATP(CTP):tRNA nucleotidyl-transferase^{13,14}, an enzyme probably encoded by the still uncharacterized NEQ152 gene. By using a RT-PCR approach involving circularization of the tRNA¹⁵ we were able to identify the 5' and 3' ends of the mature tRNA. Our sequencing results show size-maturation of the joined tRNA^{Glu}, as a CCA sequence is indeed added to the 3' end of both tRNA^{Glu} isoacceptors after transcription (Fig. 2b). A final requirement for tRNA functionality *in vivo* is the ability to serve as a substrate for amino acid attachment by

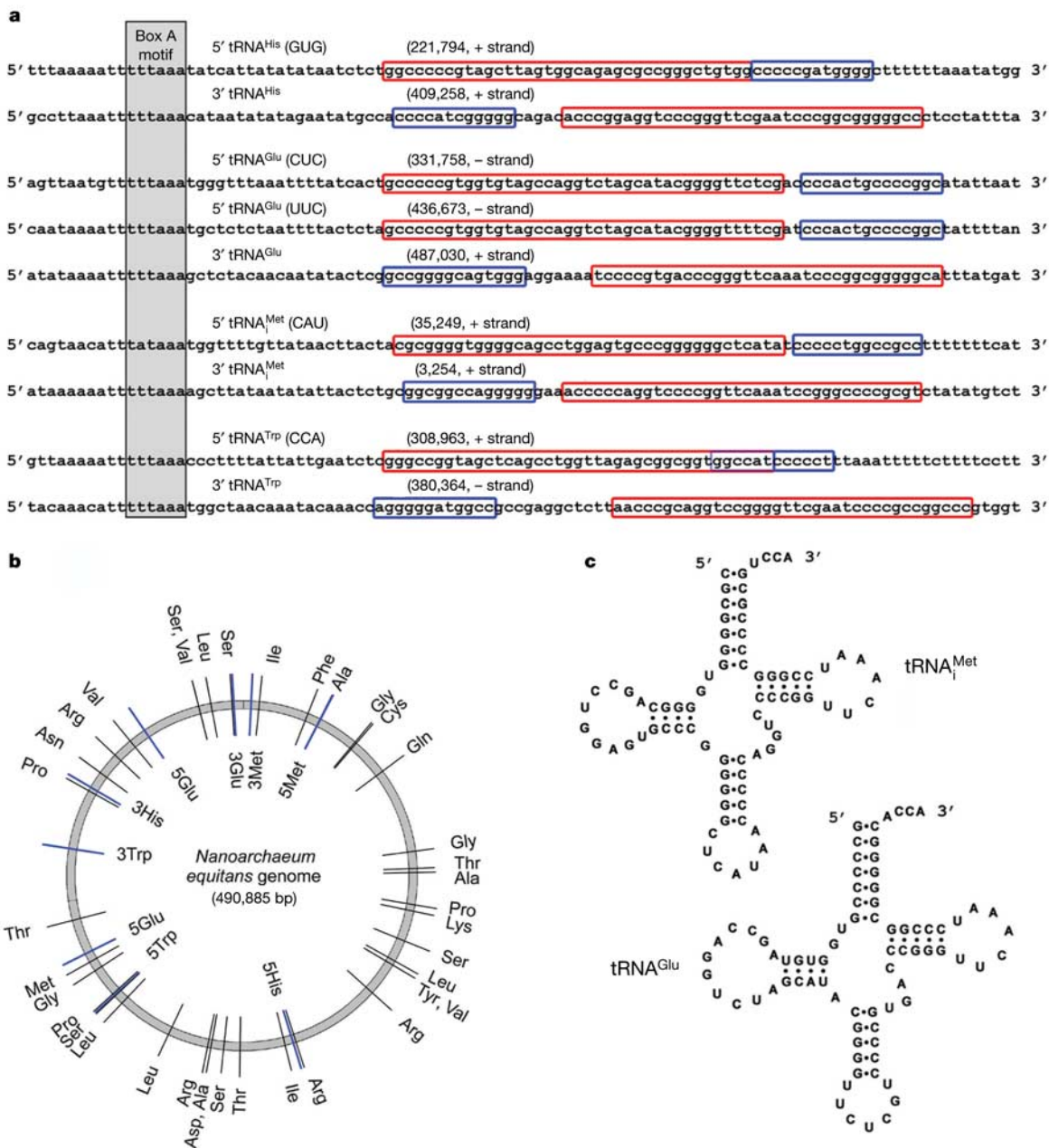


Figure 1 Predicted split *N. equitans* tRNA genes. **a**, tRNA half genes. The archaeal RNA polymerase III promoter consensus box A motif, the tRNA half genes (red) and intervening reverse complementary sequences (blue) are indicated. The positions of the tRNA halves in the *N. equitans* genome and the strand are indicated. **b**, Schematic

representation of the genomic distribution of tRNA genes (indicated by the amino-acid three-letter code) and tRNA half genes (5 indicates the 5' tRNA half gene, 3 indicates the 3' tRNA half gene) identified by our search algorithm. **c**, The joined sequences of tRNA^{Glu} and tRNA^{Met} as verified by RT-PCR and sequencing (see Fig. 2).

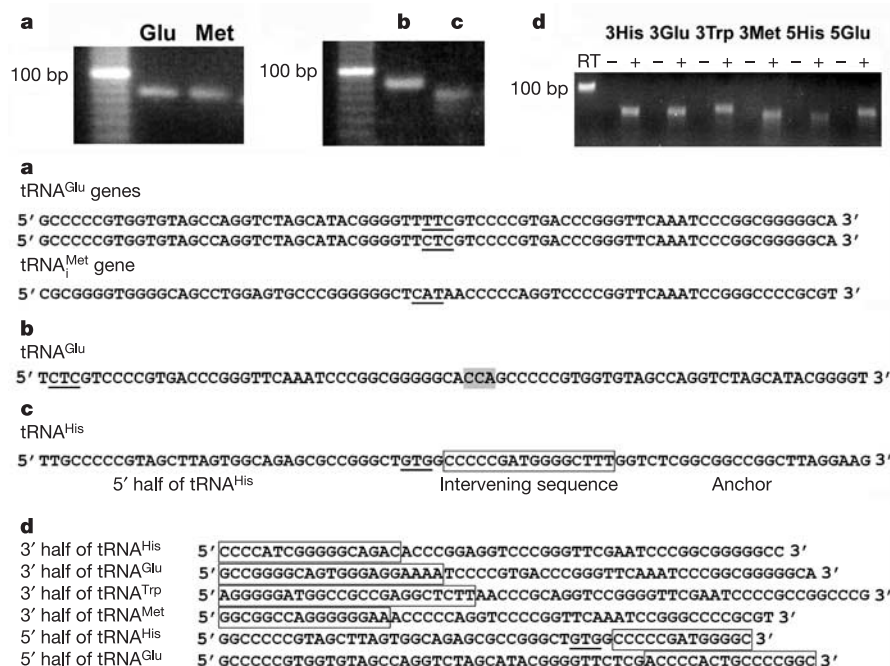


Figure 2 RT-PCR amplification of the newly identified tRNAs in *N. equitans*. The RT-PCR products were separated on a 4% ethidium bromide-stained acrylamide gel and compared with a 10-bp ladder marker (the 100-bp fragment is indicated). For RT-PCR of the tRNA halves, control experiments without reverse transcriptase (RT —) were performed to ensure that solely RNA transcripts were amplified (RT +). The different RT-PCR techniques employed are detailed in Methods. Sequencing results are displayed

with the anticodon underlined. The post-transcriptionally edited CCA terminus is shaded. The intervening complementary regions of the primary tRNA half transcripts are boxed. **a**, Detection of full-length tRNA by RT-PCR; **b**, detection of 3' maturation (CCA addition) by RT-PCR of circularized tRNA; **c**, detection of primary transcripts by RT-PCR of anchor-ligated tRNA; **d**, detection of primary transcripts by RT-PCR.

aminoacyl-tRNA synthetases. Aminoacylation reactions were performed to verify acceptor activity of the joined tRNAs. For this reason, the *N. equitans* genes encoding histidyl-tRNA synthetase (HisRS) and glutamyl-tRNA synthetase were cloned. The two enzymes were produced in *Escherichia coli*, and HisRS was purified by flocculation at 80 °C. Both synthetases were active and were able to acylate total *N. equitans* tRNA with their cognate ¹⁴C-labelled amino acids (Fig. 3). Direct proof of the identity of the aminoacylated tRNA was obtained by northern blot analysis of acid/urea gels¹⁶ after the separation of Glu-tRNA and His-tRNA from deacylated tRNA due to a difference in electrophoretic mobility between the two species (Fig. 3). The oligonucleotide probes for hybridization were complementary to a region comprising the anticodon stem/loop of the full-length tRNA^{His} and tRNA^{Glu}. These results strongly indicate an active role of these mature joined tRNAs in protein biosynthesis.

The 'missing' *N. equitans* tRNAs identified here reveal the necessity for assembly of two tRNA half molecules. What is the mechanism of joining these tRNA halves? We propose a model based on the discovery of extended reverse complementary intervening sequences (Fig. 4). Earlier studies showed that a GC-rich intervening RNA duplex increases the efficiency of intermolecular splicing (*trans*-splicing) of mRNA precursors *in vitro*^{17,18}. *Trans*-splicing *in vivo* occurs in several plant mitochondrial transcripts encoding subunits of the NADH dehydrogenase complex. In most cases the exon-flanking regions form a group II intron structure. An extended GC-rich duplex in the split intron is thought to facilitate base pairing of the two intron halves¹⁹. Similarly, during *trans*-splicing of the *N. equitans* tRNA halves a 12–14-bp fully paired RNA duplex in the intervening sequences would be the primary nucleation region in annealing the two RNA sequences. This duplex would

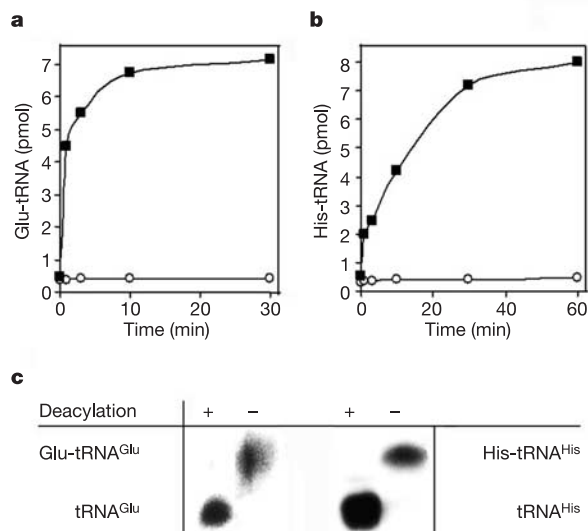


Figure 3 Aminoacylation of unfractionated *N. equitans* tRNA by *N. equitans* glutamyl-tRNA synthetase and histidyl-tRNA synthetase. The purification and aminoacylation procedures are detailed in Methods. **a**, Glutamylation with *N. equitans* glutamyl-tRNA synthetase (filled squares) at 37 °C. Open circles, control reactions without tRNA. **b**, Histidylation with *N. equitans* histidyl-tRNA synthetase (filled squares) at 80 °C. Open circles, control reactions without tRNA. **c**, Northern blot analysis of *N. equitans* tRNA glutamylated by glutamyl-tRNA synthetase or histidylated by HisRS for 1 h reaction time. The resulting aminoacyl-tRNA was loaded onto an acid/urea gel directly (—) or after deacylation (+). The blots were probed with a ³²P-labelled oligonucleotide spanning the joined anticodon stem/loop.

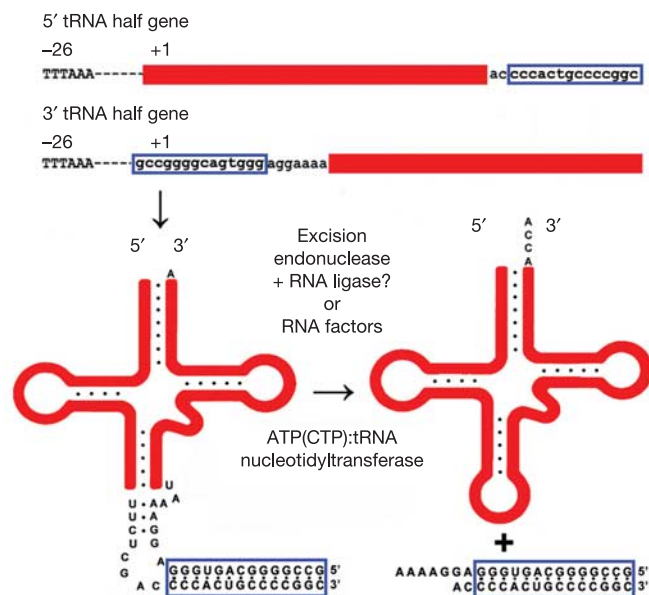


Figure 4 Schematic representation of a 5' tRNA half gene (tRNA^{Glu}) and the corresponding 3' tRNA half gene found in *N. equitans*. The archaeal RNA polymerase III promoter consensus sequence (TTTAAA), the tRNA half genes (red) and the intervening reverse complementary sequences that are supposed to facilitate joining of the halves (blue) are indicated.

facilitate folding of the whole tRNA body and stabilize the cloverleaf structure of the tRNA. It should be noted that for the joined tRNA^{His} and tRNA^{Glu} the region between the folded tRNA and the intervening RNA duplex resembles the consensus bulge–helix–bulge motif structure described for archaeal and eukaryotic tRNA introns^{20,21}. Because this structure is located at the position where most archaeal tRNA introns occur, a similar mechanism of tRNA maturation is possible. In this case one of the two putative tRNA splicing endonucleases (NEQ205 and NEQ261)^{22,23} might be responsible for intermolecular tRNA splicing, and an RNA ligase would join the 5' and 3' tRNA half molecules²⁴. A second possibility is an RNA-mediated *trans*-splicing mechanism. Both possibilities will be investigated.

Why does *N. equitans* employ this strategy? The advantage of tRNA *trans*-splicing is not apparent, given the small size of a tRNA gene. What is remarkable is the finding that one single exon (3' tRNA^{Glu} half) is *trans*-spliced to two exons (5' tRNA^{Glu} half and 5' tRNA^{Glu} half). It has been suggested⁴ that in the pre-biotic world two RNA hairpins had the simplest RNA structure and folded by given similarity into a cloverleaf-like structure. After the birth of the cloverleaf shape some template RNAs would evolve into ancient tRNAs. The intervening complementary sequences of the *N. equitans* split tRNAs might indicate an intermediate state in which the hairpins are still separated and have to be joined and stabilized by a GC-rich duplex. It is possible that the 3' tRNA half comprised of a highly conserved T-loop minigene could be fused to various anticodon-containing 5' tRNA halves to satisfy the growing complexity of protein translation during evolution^{25,26}. Thus, introns in modern tRNAs might be remnants of this duplex from an earlier world, which still performs its function in *N. equitans*.

Is this the only example of split genes for stable RNAs? Although we do not know of any other occurrences, it should be noted that the *N. equitans* genome does not possess an orthologue of the RNA component of RNase P²⁷ nor any recognizable fragment genes. It remains to be determined whether the *N. equitans* tRNA transcripts

contain leader sequences upstream of the 5' end of tRNA. In their absence, there would not be a need for the presence of RNase P in the organism.

An extensive search of the available bacterial and archaeal genome sequences did not reveal split tRNA genes in other organisms. Future sequences of other very small or compact genomes should reveal whether split tRNA genes are signs of a very early genome² or whether they are created in a later process of genome size reduction. The sequencing of other Nanoarchaeota genomes is therefore eagerly awaited. □

Methods

Computational method for tRNA identification

tRNA genes were predicted by use of a new bioinformatics approach and the program Virtual Footprint (<http://www.prodoric.de/sts/>). Position weight matrices were generated from both a conserved, continuous 3' region of tRNA genes (nt 54–76) and a 5' region of tRNA genes (nt 1–16) in an alignment of more than 4,000 tRNA genes (taken from ref. 6). For this purpose the information content was used as scoring function²⁸ with slight modifications. tRNA gene searches were performed with these position weight matrices on a genome scale with the highest sensitivity (the threshold score was taken from the lowest scoring sequence of the training set). The information that the 3' region contains a pairing stretch of 7 nt to a reverse complementary part in the 5' region (the tRNA's acceptor stem) was used to identify matching pairs of tRNA gene halves. Using this approach, all the previously annotated tRNAs were identified, including nine additional tRNA halves that fell into the threshold range of the annotated tRNAs.

Cell culture and tRNA isolation

N. equitans cells were grown in a 300-l fermenter in a simultaneous culture with *Ignicoccus* sp. and purified by gradient centrifugation as described¹. The cell pellet was lysed by chemical digestion with 2% SDS, followed by the isolation and purification of total RNA as described²⁹. The tRNA was further purified by MonoQ HR 5/5 anion-exchange chromatography to eliminate residual genomic DNA contamination. The tRNA was eluted with a linear 60-ml gradient of 0–1 M NaCl in 20 mM MOPS pH 6.2.

Reverse transcription and sequencing

Total tRNA from *N. equitans* was reverse transcribed with Thermoscript reverse transcriptase and PCR amplified with Platinum Taq DNA polymerase (Invitrogen) according to the manufacturer's directions. The tRNA was denatured at 100 °C for 5 min and snap-cooled on ice for 5 min to facilitate transcription through the highly stable secondary structure of the tRNA. PCR products were cloned with the pCR-2.1-TOPO cloning kit (Invitrogen). Plasmids were sequenced at the W. M. Keck Facility. The following oligonucleotides were used for PCR amplification of the indicated full-length tRNAs: 5'-TGCCCCCGCCGATTGAACC-3' and 5'-GCCCGGTGGTGTAGCCAGG TCTAGC-3' (tRNA^{Glu} (UUC), tRNA^{Glu} (CUC)); 5'-ACGCGGGGCCCGATTGAACC-3' and 5'-CGCGGGGTGGGCGAGCTGGAGTGC-3' (tRNA^{Met}). The reverse primers for RT-PCR of the 3' tRNA halves (Fig. 2d) comprised 22 nt, whereas the forward primers were 17 nt (tRNA^{His}, tRNA^{Glu}) or 18 nt (tRNA^{Trp}, tRNA^{Met}) long. The reverse primers for RT-PCR of the 5' tRNA halves comprised 26 nt and the forward primers were 17 nt (tRNA^{His}) or 19 nt (tRNA^{Glu}) long.

A different RT-PCR approach uses total *N. equitans* tRNA circularized by RNA-ligase-mediated joining of the 5' and 3' ends of a tRNA as described previously¹⁵. Circularized tRNA^{Glu} was PCR amplified, cloned and sequenced as described above by using the oligonucleotides 5'-CGAGAACCCTGATGCTAGACCTGGCTACAC-3' and 5'-TCTCGTCCCCGTGACCCGGTTCAAATCCC-3'. In a third RT-PCR approach an anchor oligonucleotide (5'-pGGTCTCGGCGGCCGCTTAGGdC-3') was ligated to total *N. equitans* tRNA as described³⁰ and cDNA clones were produced and sequenced as described above. Two oligonucleotides were used for PCR amplification of the anchored 5' tRNA^{His} half: 5'-GCCCGGTAGCTTAGTGGCAGAG-3' and 5'-CCTAAGCC GGCCGCCGAGACC-3'.

Preparation of proteins and aminoacylation assay

N. equitans *hisS* (NEQ102) and *gltX* (NEQ302) genes were amplified by PCR from genomic DNA. The genes were cloned into the *NdeI/EcoRI* site of pET12b(+) (Invitrogen) for *hisS* and into the *EcoRV* site of petBlue (Invitrogen) for *gltX*, to facilitate expression of the proteins in the *E. coli* BL21-codon plus (DE3)-RIL strain (Stratagene). Cultures were grown at 37 °C in Luria–Bertani medium supplemented with 100 µg ml⁻¹ ampicillin and 34 µg ml⁻¹ chloramphenicol. Expression of the recombinant proteins was induced for 4 h at 37 °C by the addition of 1 mM isopropyl α-D-thiogalactopyranoside before cell harvesting. Cells were resuspended in buffer containing 50 mM Tris–HCl pH 7.5 and 300 mM NaCl, and broken by sonication. The fractions were extensively flocculated at 80 °C for 30 min, then centrifuged for 30 min at 20,000 g. Aminoacylation was performed in a 0.1 ml reaction at 80 °C in 50 mM Hepes pH 7.0, 50 mM KCl, 4 mM ATP, 15 mM MgCl₂, 3 mM dithiothreitol, 5 µg total *N. equitans* tRNA, 50 µM [¹⁴C]glutamate (256 mCi mmol⁻¹; 9.47 GBq mmol⁻¹) or 50 µM [¹⁴C]histidine (314 mCi mmol⁻¹; 11.6 GBq mmol⁻¹), and the aminoacyl-tRNA synthetase. Glutamyl-tRNA synthetase activity assays revealed advanced activity at 37 °C reaction temperature. Aliquots of 20 µl were removed at the intervals indicated in Fig. 3, and radioactivity was measured as described²⁹. Separation of tRNA by acid/urea gel electrophoresis (9.6% gel run for 40 h (tRNA^{Glu}) and 6.4% gel run for 22 h (tRNA^{His})) and electroblotting onto Hybond N⁺

membrane (Amersham Biosciences) were performed as described¹⁶. Northern analysis was performed with ³²P-labelled oligonucleotides complementary to bases 12–50 of *N. equitans* tRNA^{Glu} and bases 11–50 of *N. equitans* tRNA^{His}.

Received 30 August; accepted 2 December 2004; doi:10.1038/nature03233.

- Huber, H. et al. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67 (2002).
- Waters, E. et al. The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl Acad. Sci. USA* **100**, 12984–12988 (2003).
- Di Giulio, M. The non-monophyletic origin of the tRNA molecule. *J. Theor. Biol.* **197**, 403–414 (1999).
- Tanaka, T. & Kikuchi, Y. Origin of the cloverleaf shape of transfer RNA—the double-hairpin model: Implication for the role of tRNA intron and the long extra loop. *Viva Origino* **29**, 134–142 (2001).
- Boucher, Y. & Doolittle, W. F. Something new under the sea. *Nature* **417**, 27–28 (2002).
- Marck, C. & Grosjean, H. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8**, 1189–1232 (2002).
- Eddy, S. R. & Durbin, R. RNA sequence analysis using covariance models. *Nucleic Acids Res.* **22**, 2079–2088 (1994).
- Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
- Marck, C. & Grosjean, H. Identification of BBH splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. *RNA* **9**, 1516–1531 (2003).
- Hain, J., Reiter, W. D., Hudepohl, U. & Zillig, W. Elements of an archaeal promoter defined by mutational analysis. *Nucleic Acids Res.* **20**, 5423–5428 (1992).
- Connolly, S. A., Rosen, A. E., Musier-Forsyth, K. & Francklyn, C. S. G-1:C73 recognition by an arginine cluster in the active site of *Escherichia coli* histidyl-tRNA synthetase. *Biochemistry* **43**, 962–969 (2004).
- Sekine, S. et al. Major identity determinants in the 'augmented D helix' of tRNA(Glu) from *Escherichia coli*. *J. Mol. Biol.* **256**, 685–700 (1996).
- Schurer, H., Schiffer, S., Marchfelder, A. & Mörl, M. This is the end: processing, editing and repair at the tRNA 3'-terminus. *Biol. Chem.* **382**, 1147–1156 (2001).
- Xiong, Y., Li, F., Wang, J., Weiner, A. M. & Steitz, T. A. Crystal structures of an archaeal class I CCA-adding enzyme and its nucleotide complexes. *Mol. Cell* **12**, 1165–1172 (2003).
- Lohan, A. J. & Gray, M. W. Methods for analysis of mitochondrial tRNA editing in *Acanthamoeba castellanii*. *Methods Mol. Biol.* **265**, 315–332 (2004).
- Varshney, U., Lee, C. P. & Rajbhandary, U. L. Direct analysis of aminoacylation levels of tRNAs *in vivo*. Application to studying recognition of *Escherichia coli* initiator tRNA mutants by glutamyl-tRNA synthetase. *J. Biol. Chem.* **266**, 24712–24718 (1991).
- Konarska, M. M., Padgett, R. A. & Sharp, P. A. Trans splicing of mRNA precursors *in vitro*. *Cell* **42**, 165–171 (1985).
- Solnick, D. Does trans splicing *in vitro* require base pairing between RNAs? *Cell* **44**, 211 (1986).
- Wissinger, B., Schuster, W. & Brennicke, A. Trans splicing in *Oenothera* mitochondria: nad1 mRNAs are edited in exon and trans-splicing group II intron sequences. *Cell* **65**, 473–482 (1991).
- Abelson, J., Trotta, C. R. & Li, H. tRNA splicing. *J. Biol. Chem.* **273**, 12685–12688 (1998).
- Fabbri, S. et al. Conservation of substrate recognition mechanisms by tRNA splicing endonucleases. *Science* **280**, 284–286 (1998).
- Li, H. & Abelson, J. Crystal structure of a dimeric archaeal splicing endonuclease. *J. Mol. Biol.* **302**, 639–648 (2000).
- Kleman-Leyer, K., Armbruster, D. W. & Daniels, C. J. Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. *Cell* **89**, 839–847 (1997).
- Salgia, S. R., Singh, S. K., Gurha, P. & Gupta, R. Two reactions of *Haloferax volcanii* RNA splicing enzymes: joining of exons and circularization of introns. *RNA* **9**, 319–330 (2003).
- Maizels, N. & Weiner, A. M. Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc. Natl Acad. Sci. USA* **91**, 6729–6734 (1994).
- Nagaswamy, U. & Fox, G. E. RNA ligation and the origin of tRNA. *Orig. Life Evol. Biosph.* **33**, 199–209 (2003).
- Gopalan, V., Vioque, A. & Altman, S. RNase P: variations and uses. *J. Biol. Chem.* **277**, 6759–6762 (2002).
- Schneider, T. D., Stormo, G. D., Gold, L. & Ehrenfeucht, A. Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* **188**, 415–431 (1986).
- Curnow, A. W., Tumbula, D. L., Pelaschier, J. T., Min, B. & Söll, D. Glutamyl-tRNA^{Gln} amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis. *Proc. Natl Acad. Sci. USA* **95**, 12838–12843 (1998).
- Williams, M. A., Johzuka, Y. & Mulligan, R. M. Addition of non-genomically encoded nucleotides to the 3'-terminus of maize mitochondrial mRNAs: truncated rps12 mRNAs frequently terminate with CCA. *Nucleic Acids Res.* **28**, 4444–4451 (2000).

Acknowledgements We thank H. Huber and K. O. Stetter for advice and spirited discussions, M. Thomm for the use of laboratory facilities, and J. Yuan and J. Sabina for critically reading the manuscript. This work was supported by grants from the National Institute of General Medical Sciences and the Department of Energy (to D.S.) and by the German Federal Ministry of Education and Research (BMBF) for the Bioinformatics Competence Center 'Intergenomics' (to D.J.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.S. (sol@trna.chem.yale.edu).

Chemical structure and biological activity of the *Caenorhabditis elegans* dauer-inducing pheromone

Pan-Young Jeong¹, Mankil Jung², Yong-Hyeon Yim⁴, Heekyeong Kim², Moonsoo Park², Eunmi Hong¹, Weontae Lee¹, Young Hwan Kim⁵, Kun Kim³ & Young-Ki Paik¹

¹Department of Biochemistry and Yonsei Proteome Research Center, ²Department of Chemistry, ³Bioproducts Research Center, Yonsei University, Seoul 120-749, Korea

⁴Korea Research Institute of Standards and Science, Taejeon 305-600, Korea

⁵Korea Basic Science Institute, Taejeon 305-333, Korea

Pheromones are cell type-specific signals used for communication between individuals of the same species. When faced with overcrowding or starvation, *Caenorhabditis elegans* secrete the pheromone daumone, which facilitates communication between individuals for adaptation to adverse environmental stimuli^{1–4}. Daumone signals *C. elegans* to enter the dauer stage, an enduring and non-ageing stage of the nematode life cycle with distinctive adaptive features and extended life. Because daumone is a key regulator of chemosensory processes in development and ageing^{5,6}, the chemical identification of daumone is important for elucidating features of the daumone-mediated signalling pathway. Here we report the isolation of natural daumone from *C. elegans* by large-scale purification, as well as the total chemical synthesis of daumone. We present the stereospecific chemical structure of purified daumone, a fatty acid derivative. We demonstrate that both natural and chemically synthesized daumones equally induce dauer larva formation in *C. elegans* (N2 strain) and certain dauer mutants, and also result in competition between food and daumone. These results should help to elucidate the daumone-mediated signalling pathway, which might in turn influence ageing and obesity research and the development of antinematodal drugs.

Although general properties of partially enriched daumone have been known for more than two decades², its biochemical identity and other chemical characteristics have not yet been determined. Because there is no chemical method for confirming the presence of very low concentrations of daumone in crude cell extracts², we used a large-scale culture (with a 300-litre fermenter), and performed a dauer formation assay with every fraction at each purification step. Following two consecutive organic solvent extractions, the ethyl acetate extracts were loaded onto a silica gel column, which was then washed once with a solution of hexane/ethyl acetate/methanol

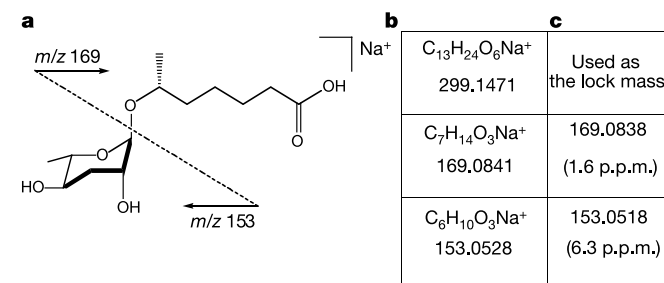


Figure 1 Molecular structure of daumone: (—) (6R)-(3,5-dihydroxy-6-methyltetrahydropyran-2-yl)oxyheptanoic acid. **a**, Structure of daumone and the proposed fragmentation pattern for sodiated daumone, using m/z 299.1471 as the lock mass for mass calibration. **b**, Elemental composition and accurate mass calculated. **c**, Accurate mass of each fragment measured.

(7:7:1, v/v/v). Fractions bound to the silica gel column were eluted with methanol, which facilitates the removal of strong non-polar components and resulted in one large activity peak. Having confirmed the ability of the eluents from the silica gel column to induce dauer formation, this fraction was further subjected to separation on amine and gel permeation chromatography (GPC) columns. This resulted in a functional daumone with >90% purity, as judged by its appearance as a dominant peak in the mass spectra obtained by electrospray ionization/mass spectrometry (ESI/MS) (Table 1). For these later stages of the purification process, representative activity peak profiles from the amine and GPC columns (see Supplementary Fig. 1a, b) were obtained; the daumone yield was approximately 22.5 mg, corresponding to about >15,000-fold purification based on specific activity (or 128,300-fold purification based on mass). Note that there is a discrepancy between the theoretical purification yield and the dauer-formation activity of daumone. There are at least two possible reasons for this phenom-

enon; first, substantial loss of daumone might occur during purification. Second, the purified or synthetic daumone might be more active when it is associated with other unidentified molecules (for example, membrane lipids); the low solubility of purified daumone in aqueous environments (such as plate agar) might result in less efficient interaction with sensory neurons on the body surface compared to its less purified forms. This remains to be established.

To determine the chemical structure of the isolated daumone, the active fraction of the GPC column was subjected to various chemical analyses including ESI/MS and NMR spectroscopy. On the basis of the elemental composition, ESI and quadrupole time-of-flight (Q-TOF2) mass spectral data, it was confirmed that the compound had a molecular mass (M_r) of 276 and the molecular formula $C_{13}H_{24}O_6$. The accurate masses obtained for $(M + Na)^+$ and $(M + NH_4)^+$, where M is the parent molecular ion mass, are 299.1480 (calculated $M_r = 299.1471$; M_r measurement error 3.1 p.p.m.) and 294.1904 (calculated $M_r = 294.1917$; M_r measure-

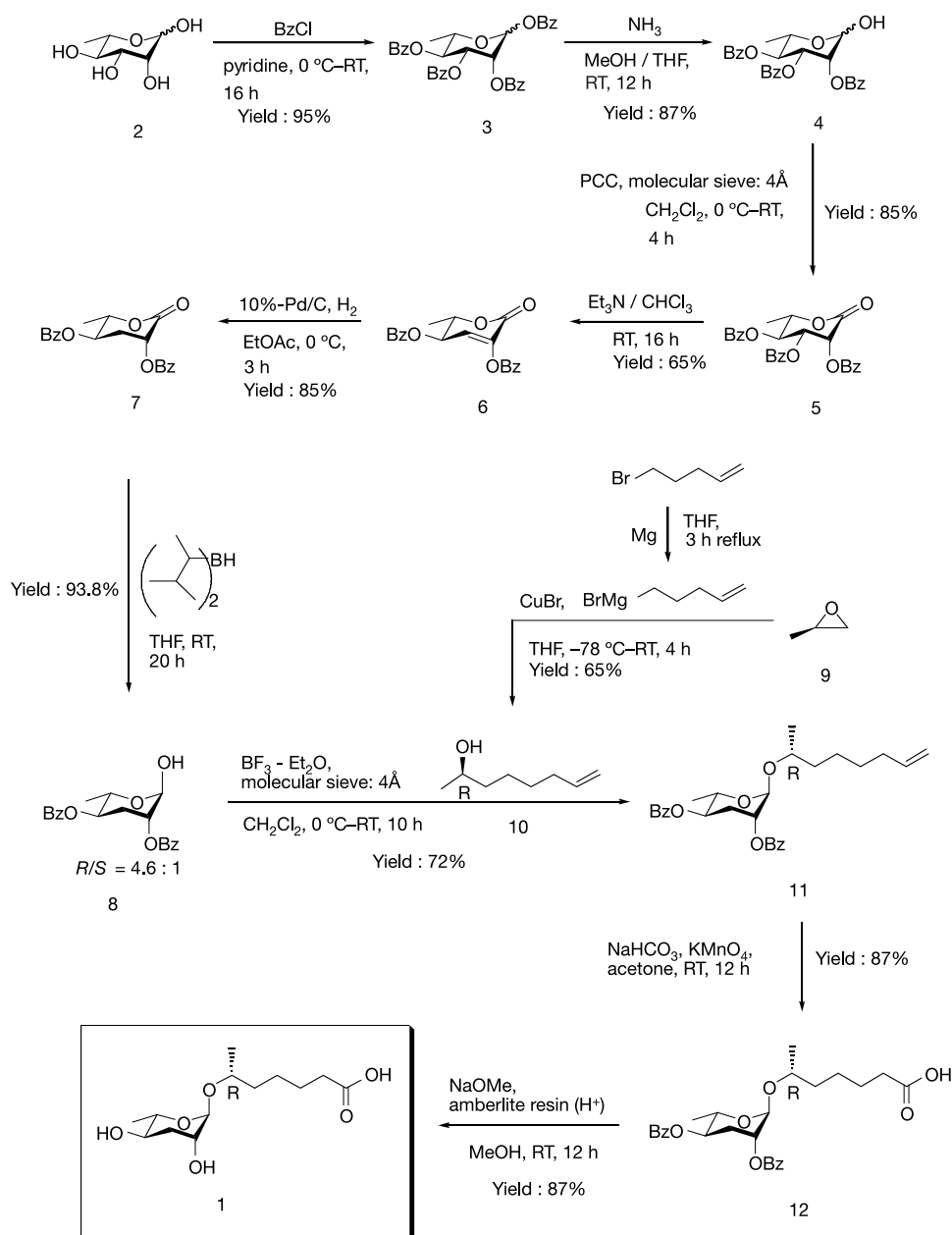


Figure 2 Stereospecific total synthesis of daumone. Synthesis of daumone (**1**) was performed in ten steps. The key intermediate (**8**) is synthesized from L-rhamnose (**2**). RT, room temperature (23–25 °C); R, S, stereochemical absolute configurations.

Table 1 Purification of daumone from a large-volume culture of *C. elegans*

Purification step	Total content (g)	Dilution for dauer assay	Aliquot size (g)	Dauer formation* (%)	Specific activity† (%)	Fold purification (specific activity)‡	Fold purification (mass)§
A. Dry broth culture	2887.5	1/1,000	2.8875	83.8	29	1	1
B. Ethanol extraction	279.0	1/1,000	0.279	92.1	330	11	10
C. Ethyl acetate extraction	60.0	1/1,000	0.06	99.6	1660	57	48
D. Silica gel column	1.65	1/100	0.0165	97.2	5891	203	1,750
E. Amine column	0.24	1/100	0.0024	95.3	39708	1,369	12,031
F. GPC column	0.0225	1/100	0.000225	98.4	437333	15,080	128,300

A representative purification procedure from a 300-litre culture of *C. elegans* is shown.

*Average dauer formation (%) is based on dauer-formation assays (performed in duplicate) using the aliquot size indicated in column 4.

†Specific activity is defined as % dauer formation per plate at 25 °C (80 µg *E. coli*) and was calculated by dividing % dauer formation (column 5) by aliquot size (column 4).

‡Purification based on specific activity is calculated by dividing the specific activity value of the fraction obtained from each purification step (column 6, rows B–F) by that of the starting material (column 6, row A).

§Purification based on mass is calculated by dividing content (g) of the starting material (column 2, row A) by the weight (yield) of the extract obtained at each purification step (column 2, rows B–F).

ment error 4.3 p.p.m.), respectively. Proton and carbon-13 NMR assignments were easily performed from spectra obtained using the combined techniques of heteronuclear multiple-quantum coherence (HMQC), distortionless enhancement by polarization transfer (DEPT), heteronuclear multiple-bond correlation (HMBC) and double quantum filtered-correlation spectroscopy (DQF-COSY) (see Supplementary Table 1). The daumone structure was proposed on the basis of the MS and NMR data (Fig. 1). Data from total correlation spectroscopy (TOCSY) and low-resolution MS (see Supplementary Fig. 1c–f) showed no correlation or linkage between the two protons located between the pyran ring and the heptanoic acid. The relative stereochemistry of the daumone isolated from *C. elegans* was unambiguously established by rotating frame nuclear Overhauser enhancement spectroscopy (ROESY), nuclear Overhauser and exchange spectroscopy (NOESY), two-dimensional (2D)-NMR and molecular modelling analysis as 6*R*, 2'*R*, 3'*R*, 5'*R*, 6'*S* (see Supplementary Figs 2–5). These data combined with ¹H- and ¹³C-NMR spectra yielded a chemical structure consistent with (–) 6-(3,5-dihydroxy-6-methyltetrahydropyran-2-yloxy) heptanoic acid (Fig. 1), establishing the identity of the pure daumone. The exact stereochemistry of daumone ((–)-6*R*-(3'*R*, 5'*R*-dihydroxy-6'*S*-methyltetrahydro-pyran-2'*R*-yloxy) heptanoic acid) was established by synthesizing the molecule chemically. Briefly, stereo-specific total synthesis of the pheromone was designed and successfully performed in ten steps, in which the key intermediate (**8**) was synthesized starting from L-rhamnose (**2**) as shown (Fig. 2). The synthetic compound (**1**) is identical to the natural pheromone isolated from *C. elegans*, according to spectroscopic data: 2D-NMR (500 MHz), ¹³C NMR, infrared spectroscopy, high-resolution mass spectrometry (HRMS) and specific rotation $[\alpha]_D^{25} = -81.0$

(concentration: 0.1 g ml^{–1}, CH₃OH), where D is the D line of the sodium spectrum. Total synthesis of compound (**1**) confirmed its absolute stereochemistry with natural pheromone as 6*R*, 2'*R*, 3'*R*, 5'*R*, 6'*S* (see Supplementary Tables 3 and 4). We propose that this compound corresponds to the first reported dauer-inducing substance and that this is the first description of the isolation and synthesis of the compound.

To determine whether both the natural and synthetic daumone induce dauer larva formation, we grew *C. elegans* at low population density in the presence of fixed concentrations of natural or synthetic daumone (384 µM) and *E. coli* (160 µg per plate) (see below) and then visualized several dauer-specific features by Nomarski microscopy. As expected, in the presence of natural daumone, worms acquire the full complement of dauer-specific morphological characteristics, including thin body shape (Fig. 3a, b), swollen hypodermal bodies (Fig. 3c), increased fat storage (Fig. 3d) and dauer-specific alae (Fig. 3e). Because identical results were obtained using either synthetic or natural daumone at the same concentrations (Fig. 3a, b), we used synthetic daumone in most experiments unless otherwise specified. Only eggs and worms at the L1 and early L2 stages sense pure daumone and form dauer larvae (at daumone concentrations of 384 µM); worms at the late L2 stage or later do not become dauer larvae, as previously reported^{3,4} (data not shown).

Owing to the availability of synthetic daumone, we first determined the quantitative relationship between two factors which have been proposed to compete with each other at dauer-permissible temperatures²: daumone and food signal. Our results show that there appears to be a quantitative relationship (that is, conditional competition) between daumone and food signal (see Supplemen-

Table 2 Effects of synthetic and natural daumone on dauer formation in nematodes

Strain	Genotype	Dauer formation (%)							
		Phenotype of progeny at 20 °C, (%)		<i>n</i>		Phenotype of progeny at 25 °C, (%)		<i>n</i>	
		S	N	S	N	S	N	S	N
<i>C. elegans</i>	N2	51 ± 2.1	42 ± 9.4	274	172	99 ± 1.6	97 ± 1.2	192	183
	<i>daf-3(mg90)</i>	52 ± 15.9	39 ± 2.3	75	93	82 ± 9.5	51 ± 6.8	148	142
	<i>daf-5(e1386)</i>	26 ± 12.3	21 ± 0.4	94	79	45 ± 2.9	45 ± 9.1	118	107
	<i>daf-10(p821)</i>	8 ± 2.1	5 ± 1.3	129	157	82 ± 3.8	79 ± 7.4	162	149
	<i>daf-11(m47)</i>	12 ± 4.9	11 ± 2.0	169	129	97 ± 1.4	92 ± 0.7	126	79
	<i>daf-12(rh284)</i>	12 ± 5.4	8 ± 2.2	157	121	92 ± 2.2	94 ± 1.8	172	153
	<i>daf-22(m130)</i>	31 ± 3.9	47 ± 9.9	129	160	47 ± 4.2	59 ± 8.0	142	131
<i>C. briggsae</i>	Wild type	69 ± 2.7	73 ± 6.1	129	163	96 ± 1.2	91 ± 1.3	124	110
<i>C. elegans</i> *		<i>daf-3(e1376)</i> , <i>daf-6(e1377)</i> , <i>daf-10(e1387)</i> , <i>daf-12(rh61)</i> , <i>daf-12(rh62rh157)</i> , <i>daf-12(m583)</i> , <i>daf-12(rh286)</i> , <i>daf-12(rh61rh412)</i> , <i>daf-12(m20)</i> , <i>daf-12(m25)</i> , <i>daf-12(m116)</i> , <i>daf-12(m421)</i> , <i>daf-13(m66)</i> , <i>daf-16(mu86)</i> , <i>daf-18(e1375)</i> , <i>daf-18(ok480)</i> , <i>che-3(e1253)</i> , <i>che-3(e1378)</i> , <i>che-3(e1379)</i> , <i>che-3(e1239)</i> , <i>che-3(hf5)</i> , <i>che-11(m162)</i> , <i>che-11(mn387)</i>							

Eggs were collected from nematodes grown on daumone-containing plates (at 20 °C) over a 6 h period and maintained at 20 °C or 25 °C. At 20 °C, phenotypes were scored twice, first at 80 h and again 92 h after egg-laying. At 25 °C, phenotypes were scored twice, first at 52 h and again 72 h after egg-laying. Values show mean (±s.d., *n* = 3) percentage of worms forming dauers across three different plates (50-mm diameter) in each of two experiments. Total number of surviving worms (*n*) at 20 °C for 92 h or at 25 °C for 72 h were counted across the three plates. The daumone-containing plate was prepared by loading 384 µM synthetic (S) or natural (N) daumone and 160 µg food. To prepare the natural and synthetic daumones, we weighed the compounds, dissolved them in ethanol and added double-distilled water (ethanol/water = 1:1). This stock solution was divided into aliquots for each experiment.

*Dauer mutants of *C. elegans* for which no dauer is formed at either 20 °C or 25 °C

tary Table 2). At a given amount of food (for example, 80 μg per plate), the dauer-formation activity of daumone increases with daumone concentration in a concentration-dependent manner. This result establishes both the quantitative relationship between daumone concentration and dauer formation (at a given amount of food), and also the appropriate concentration of daumone for standard dauer-formation assays. On the basis of this observation and the estimated concentration of daumone in the active fraction of GPC column eluent that exhibits >98% dauer formation (a 225 μg aliquot from the GPC column corresponds to 243 μM daumone at 90% purity; see column 4, row F in Table 1), we chose to use 384 μM daumone for all experiments.

Second, dauer-formation assays were performed to examine how a variety of previously known dauer mutants respond to daumone; the results might indicate the potential target site for daumone action in the signalling pathway. As expected, most of the dauer-defective mutants or dauer-constitutive mutants showed predicted behaviours when exposed to 384 μM of either natural or synthetic

daumone (and 160 μg of food). To our surprise, some dauer-defective mutants responded to pure daumone and became dauer larvae (Table 2). For example, *daf-3(mg90)*, *daf-5(e1386)* and *daf-10(p821)*, typical dauer-deficient mutants^{7–10}, became dauer larvae in response to daumone. Two other dauer mutants, *daf-12(rh284)* (ref. 11) and *daf-22*, also formed dauer larvae, suggesting that their sensory function might be intact. Other dauer mutants, *daf-16(mu86)*, *daf-18(ok480)* and *daf-18(e1375)*, which are associated with an insulin receptor-like signalling pathway, were not responsive to daumone. In addition, *C. briggsae* also formed dauer larvae in response to synthetic daumone (69% at 20 °C, $n = 129$; 96% at 25 °C, $n = 124$). In most instances, synthetic daumone results in a slightly higher percentage of dauer formation than the natural molecule; however, natural daumone results in a higher percentage of dauer formation than synthetic daumone for the *daf-22* mutants tested. At present, we hypothesize there may be several dauer signalling pathways responding to daumone in different mutants; for example, *daf-22* mutants might use a different signalling pathway compared with other dauer mutants.

The present work describes the biochemical identification and characterization of daumone, which should provide the basis for further studies of daumone-mediated signalling pathways, ageing, obesity and the development of antineoplastic compounds¹². The next challenge will be the elucidation of the daumone metabolic pathway. The critical step of daumone biosynthesis might be one of two possible reactions: first, the condensation of a *de novo* synthesized short-chain fatty acid to a rhamnose ring by TDP- or UDP-glucuronosyltransferase, or second, modifications to the rhamnose ring and its condensation with breakdown products of long-chain fatty acids. We are currently investigating these possibilities using various mutants. Although there has been debate about whether daumone in *C. elegans* acts as a signal or as a crowd cue¹³, our data lead us to believe that daumone fits the definition of a signal molecule. Furthermore, daumone is structurally similar to the fatty acid precursors of many insect pheromones such as the fatty acid pheromone in bees¹⁴. In this regard, identification of a daumone receptor becomes very important, as it may be specifically expressed only in sensory neurons. Because our preliminary data indicate that daumone-mediated signalling is not directly related to the insulin-signalling pathway, a functional link between daumone- and transforming growth factor-beta (TGF- β)-signalling¹⁵ deserves further study. □

Methods

Strains and genetics

Wild-type nematodes were the *C. elegans* Bristol variety, strain N2, grown on nematode growth media (NGM) agar plates with *E. coli* (OP50) under standard uncrowded and well fed conditions at 20 °C unless otherwise stated^{16,17}. *C. elegans* mutant strains were obtained from the Caenorhabditis Genetics Center.

Purification of daumone

For the preparation of crude pheromone by solvent extraction, *C. elegans* were grown on S basal liquid medium in a large-scale fermenter (for example, 300 litres) for about 20 days at 20 °C with *E. coli* (OP50). In one representative experiment, culture broth was obtained by centrifugation and filtration after >70% of the worms had entered dauer stage. Filtered broth was evaporated to obtain dried broth culture (~2,887.5 g), which was then subjected to repeated extraction with ethanol and then ethyl acetate. The ethyl acetate extracts (~60 g) were then chromatographed on a silica gel column (100 $\times$ 10 cm; Merck) previously equilibrated with hexane/ethyl acetate/methanol (7:7:1); the bound fractions were eluted with methanol and further purified on an amine column (Spherisorb S10 NH₂ column, 250 $\times$ 10 mm; Waters) using gradient elution (from isopropanol to distilled water). The partially purified fraction (0.24 g) from this step was subjected to a GPC column (JAIGEL W-251, 500 mm length $\times$ 20 mm diameter; Japan Analytical Industry Co.) equilibrated with methanol. The final active fraction (~22.5 mg) was collected and used for dauer-formation assays and structural analysis. At every step, dauer formation activity was examined as described (see below)⁴.

Dauer-formation assays and Sudan-black staining

Plates (3 ml agar) were prepared with natural daumone (partially purified or final fraction, >90% pure) or synthetic daumone; an aliquot of the ethanol-dissolved daumone fraction (Table 1) was loaded onto the plate and left to dry on a clean bench¹⁸. Food was provided as

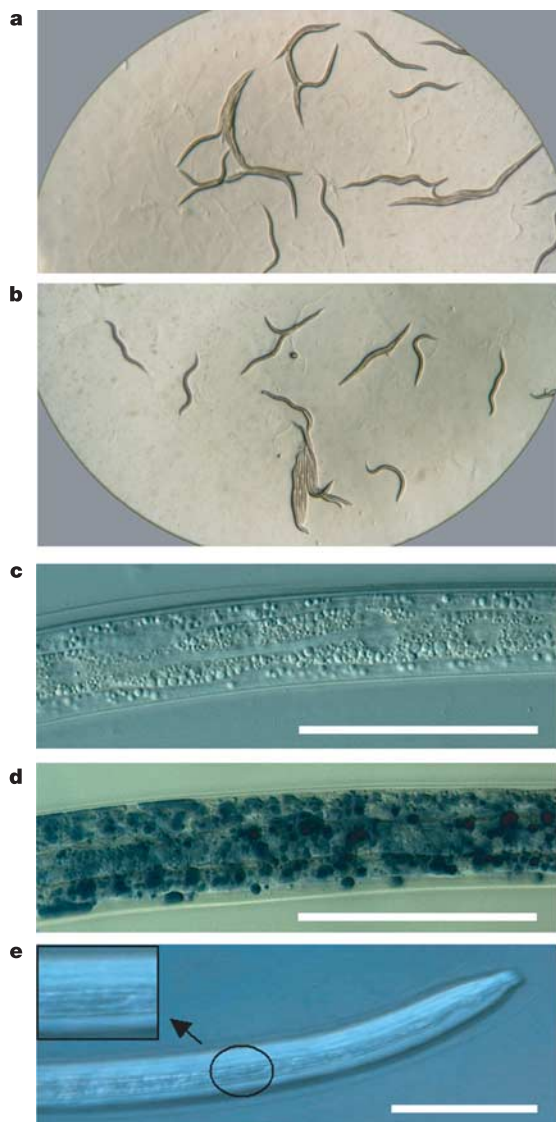


Figure 3 Validation of daumone function. **a–e**, Responses of *C. elegans* to purified natural daumone (**a**) and synthetic daumone (**b–e**). **a, b**, Micrograph illustrating thin body shape in response to natural or synthetic daumone. **c**, Nomarski photograph showing swollen, hypodermal bodies. **d**, Abundant fat storage visualized with Sudan-black staining. **e**, Dauer-specific alae. Worms were grown on plates containing 384 μM daumone and 160 μg dead *E. coli* as a food source. Scale bar, 50 μm .

a suspension of dead *E. coli* (OP50) placed on the agar medium. Briefly, daumone-containing plates (5 mm diameter) were prepared with 3 ml NGM (without peptone) as described, and known amounts of *E. coli* suspension that had been heated at 95 °C for 30 min (with vigorous vortexing every 5 min) were added. Separately, a daumone stock solution was prepared by dissolving daumone in ethanol (320 µg in 150 µl) then adding 150 µl double-distilled water to make a final solution of 320 µg in 300 µl. This daumone solution was evenly spread over the plate and allowed to dry in a clean hood. Once plates were dry, dead *E. coli* were loaded onto the center of the plate at a concentration of 160 µg per 20 µl. Plates were left to dry for an hour, then stored at 4 °C until use for dauer-formation assays.

For the dauer-formation assay, five adult *C. elegans* were placed on a plate and incubated at 20 °C for 4–6 h. After incubation, adult worms were removed and the eggs on the plate (generally 50–100 eggs per plate) were incubated for 52–72 h and then counted. A 1% SDS solution (1.0 ml) was added to the plate and the surviving worms were counted as dauer larvae. For Sudan-black staining of fat storage in the dauer larvae, worms were fixed with 1% paraformaldehyde, washed, dehydrated with ethanol and stained with dye¹⁹.

NMR analysis

For NMR experiments, the compound (~1 mg) was dissolved in deuterated methanol at a concentration of 30 mg ml⁻¹, and spectra were acquired at 25 °C with a Bruker DRX 500 MHz spectrometer. In the 2D-NMR experiments, the ¹H chemical shifts were referenced to internal sodium 4,4-dimethyl-4-silapentane-1-sulphonate (DSS). For ¹H-¹H 2D-DQF-COSY²⁰, ¹H-¹H 2D-TOCSY and ¹H-¹H 2D-ROESY experiments²¹, data were collected using a 7,002.80 Hz spectral width, 2,048 complex points in *t*₂ and 128 increments in *t*₁. ¹³C-HMQC^{22,23} and ¹³C-HMBC²⁴ spectra were recorded for the assignment of carbon and proton resonances²⁵. Proton and carbon resonances of daumone were easily assigned using the combined techniques of DEPT, HMQC, HMBC, and DQF-COSY NMR (see Supplementary Information).

Total chemical synthesis and stereospecific structure determination

A detailed description of the chemical synthesis process is described in the Supplementary Information. The synthesized compound was identified as a synthetic daumone using the established dauer-formation assay described above. The stereospecific structure of the compound was also confirmed to be identical to the natural daumone. Please refer to the Supplementary Information for experimental procedures regarding the synthesis of daumone and its intermediates (3, 4, 5, 8, 10, 11, 12 and 1), spectral data for the synthetic daumone (including ¹H NMR, ¹³C NMR, Fourier transform infrared spectroscopy (FTIR), HRMS (fast atom bombardment), DEPT, HMBC, NOESY, ROESY, high-resolution Q-TOF mass spectrometry and retention factor (*R*_f)) and elemental analysis data for the synthetic daumone.

Received 4 October; accepted 8 November 2004; doi:10.1038/nature03201.

1. Cassada, R. C. & Russell, R. L. The dauer larva, a post-embryonic developmental variant of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **46**, 326–342 (1975).
2. Golden, J. W. & Riddle, D. L. A pheromone influences larval development in the nematode *Caenorhabditis elegans*. *Science* **218**, 578–580 (1982).
3. Golden, J. W. & Riddle, D. L. The *Caenorhabditis elegans* dauer larva: developmental effects of pheromone, food, and temperature. *Dev. Biol.* **102**, 368–378 (1984).
4. Golden, J. W. & Riddle, D. L. A *Caenorhabditis elegans* dauer-inducing pheromone and an antagonistic component of the food supply. *J. Chem. Ecol.* **10**, 1265–1280 (1984).
5. Bargmann, C. I. & Horvitz, H. R. Control of larval development by chemosensory neurons in *Caenorhabditis elegans*. *Science* **251**, 1243–1246 (1991).
6. Schackwitz, W. S., Inoue, T. & Thomas, J. H. Chemosensory neurons function in parallel to mediate a pheromone response in *C. elegans*. *Neuron* **17**, 719–728 (1996).
7. Thomas, J. H., Birnby, D. A. & Vowles, J. J. Evidence for parallel processing of sensory information controlling dauer formation in *Caenorhabditis elegans*. *Genetics* **134**, 1105–1117 (1993).
8. Patterson, G. I., Kowek, A., Wong, A., Liu, Y. & Ruvkun, G. The DAF-3 Smad protein antagonizes TGF-beta-related receptor signaling in the *Caenorhabditis elegans* dauer pathway. *Genes Dev.* **11**, 2679–2690 (1997).
9. Apfeld, J. & Kenyon, C. Regulation of lifespan by sensory perception in *Caenorhabditis elegans*. *Nature* **402**, 804–809 (1999).
10. Perkins, L. A., Hedgecock, E. M., Thomson, J. N. & Culotti, J. G. Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol.* **117**, 456–487 (1986).
11. Antebi, A., Culotti, J. G. & Hedgecock, E. M. daf-12 regulates developmental age and the dauer alternative in *Caenorhabditis elegans*. *Development* **125**, 1191–1205 (1998).
12. Bird, D. M. & Opperman, C. H. *Caenorhabditis elegans*: A genetic guide to parasitic nematode biology. *J. Nematol.* **30**, 299–308 (1998).
13. Viney, M. F. & Franks, N. R. Is dauer pheromone of *Caenorhabditis elegans* really a pheromone? *Naturwissenschaften* **91**, 123–124 (2004).
14. Regnier, F. E. & Law, J. H. Insect pheromone. *J. Lipid Res.* **9**, 541–551 (1968).
15. Koga, M., Takeuchi, M., Tameishi, T. & Ohshima, Y. Control of DAF-7 TGF-α expression and neuronal process development by a receptor tyrosine kinase KIN-8 in *Caenorhabditis elegans*. *Development* **126**, 5387–5398 (1999).
16. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
17. Choi, B. K., Chitwood, D. J. & Paik, Y.-K. Proteomic changes during disturbance of cholesterol metabolism by azacoprostan treatment in *Caenorhabditis elegans*. *Mol. Cell. Proteomics* **2**, 1086–1095 (2003).
18. Vowles, J. J. & Thomas, J. H. Multiple chemosensory defects in daf-11 and daf-21 mutants of *Caenorhabditis elegans*. *Genetics* **138**, 303–316 (1994).
19. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. daf-2, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277**, 942–946 (1997).
20. Rance, M. *et al.* Improved spectral resolution in cosy 1H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**, 479–485 (1983).
21. Bax, A. & Subramanian, S. Sensitivity-enhanced two-dimensional heteronuclear shift correlation

NMR spectroscopy. *J. Magn. Reson.* **67**, 565–570 (1986).

22. Bax, A. & Davis, D. G. Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Magn. Reson.* **63**, 207–213 (1985).
23. Bax, A., Griffey, R. H. & Hawkins, B. L. Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR. *J. Magn. Reson.* **55**, 301–315 (1983).
24. Bax, A. & Summers, M. F. ¹H and ¹³C assignments from sensitivity enhanced detection of heteronuclear multiple-bond connectivity by two-dimensional multiple quantum NMR. *J. Am. Chem. Soc.* **108**, 2093–2094 (1986).
25. Schleucher, J. *et al.* A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR* **4**, 301–306 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This study was supported by a grant to Y.K.P. from the Korea Health 21 R&D Project, Ministry of Health and Welfare, Republic of Korea. We thank J.-M. Kim at KDR Biotech Co. for his support on this project, D.J. Chitwood at the USDA-ARS Nematology Lab for his critical reading and suggestions, R. Moyer at King College (USA) for editorial assistance, J. Lee at Seoul National University for discussions and the *Caenorhabditis* Genetics Center for kind provision of the *C. elegans* strains used in this study. Technical support from the LG Chem Research Center (Taejeon, Korea) was appreciated.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.K.P. (paiky@yonsei.ac.kr).

Structure and different conformational states of native AMPA receptor complexes

Terunaga Nakagawa¹, Yifan Cheng², Elizabeth Ramm¹, Morgan Sheng¹ & Thomas Walz²

¹The Picower Center for Learning and Memory, RIKEN-MIT Neuroscience Research Center, Howard Hughes Medical Institute, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

²Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA

Ionotropic glutamate receptors mediate fast excitatory synaptic transmission in the central nervous system^{1,2}. Their modulation is believed to affect learning and memory, and their dysfunction has been implicated in the pathogenesis of neurological and psychiatric diseases^{1,2}. Despite a wealth of functional data, little is known about the intact, three-dimensional structure of these ligand-gated ion channels. Here, we present the structure of native AMPA receptors (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid; AMPA-Rs) purified from rat brain, as determined by single-particle electron microscopy. Unlike the homotetrameric recombinant GluR2 (ref. 3), the native heterotetrameric AMPA-R adopted various conformations, which reflect primarily a variable separation of the two dimeric extracellular amino-terminal domains. Members of the stargazin/TARP family of transmembrane proteins co-purified with AMPA-Rs and contributed to the density representing the transmembrane region of the complex. Glutamate and cyclothiazide markedly altered the conformational equilibrium of the channel complex, suggesting that desensitization is related to separation of the N-terminal domains. These data provide a glimpse of the conformational changes of an important ligand-gated ion channel of the brain.

Ionotropic glutamate receptors include the NMDA (*N*-methyl-D-aspartate)-, AMPA- and kainate-preferring receptor channels. AMPA-Rs conduct most of the fast excitatory postsynaptic current (EPSC), and their regulation is critical for synaptic plasticity¹. GluR1–4 (also known as GluR-A–D) encode subunits of

mammalian AMPA-Rs and have similar primary structures^{4,5}.

AMPA-R subunits consist of an N-terminal domain (NTD), a ligand-binding domain (LBD), a transmembrane domain (TMD) and a short, intracellular carboxy-terminal tail (Fig. 1a). The channel-forming TMD contains three membrane-spanning segments (M1, M3, M4) and the M2 "re-entrant loop"². AMPA-Rs have been proposed to have a heterotetrameric, dimer-of-dimers organization^{6–12}; however, the complex membrane topology and the heteromeric nature of AMPA-Rs hinder accurate structural modelling of the entire receptor^{13–15}.

To investigate its intact structure, native AMPA-Rs were purified from CHAPS-solubilized rat brain synaptosomes in the presence of the antagonist NBQX (2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[f]quinoxaline-7-sulphonamide; added to block the effect on AMPA-R of endogenous glutamate, and removed during the final gel filtration chromatography step; Supplementary Fig. 1a). In negative stain electron microscopy images, the AMPA-Rs were monodispersed and homogeneous in size, but had various shapes (Fig. 1b). The AMPA-R particles were further analysed by multivariate statistical analysis, classification and multi-reference alignment. The most common class averages (Fig. 1b, insets 1–3) show a strong but featureless oblong density in the bottom, with two elongated bipartite densities asymmetrically stacked on top of it. An additional, less-well-defined density lies between the oblong domain and the two bipartite densities.

Antigen-binding fragments (Fab) that recognize the C terminus of GluR1 or GluR2 decorated the lower part of the large oblong density (Fig. 1c, e). Fab fragments that recognize the NTD of GluR1 consistently bound to the two elongated bipartite densities, which we interpret as NTD dimers (Fig. 1d). Because the C-terminal tail is small and extends from M4, the Fab labelling implies that the oblong density represents the TMD (Fig. 1a, f). The weaker density between the TMD and the NTDs probably corresponds to the LBDs (Fig. 1a, f).

Other class averages showed a 'V'-shaped conformation with a pseudo-two-fold symmetry (Fig. 1b, insets 4 and 5). In these particles the two NTD densities are separated and positioned at varying angles relative to the TMD. Recognition by all three Fab fragments suggests that these V-shaped particles are also AMPA-Rs (Fig. 1c–e, far-right panels).

To calculate three-dimensional maps we prepared AMPA-Rs by cryo-negative staining¹⁶ and recorded image pairs of specimens tilted to 50° and 0°. The images of the untilted specimens were used to classify the particles according to their shape, and the images of the tilted specimens were then used to calculate three-dimensional reconstructions of individual classes using the random conical tilt approach¹⁷.

Density maps of three different classes of particles at a resolution of ~40 Å reveal distinct conformations of the AMPA-Rs (Fig. 2a, b). Re-projections from the density maps are almost identical to the class averages, demonstrating the consistency of the three-dimensional reconstructions with the projections (Fig. 2a, b, compare left and right panels; see also Supplementary Figs 2 and 3). Statistics for the three-dimensional reconstructions are summarized in Table 1. Depending on whether the NTDs are partially overlapping (Fig. 2a) or separated (Fig. 2b), we will refer to AMPA-Rs as adopting either a type I or type II conformation, respectively.

The NTDs are homologous to the bacterial periplasmic amino-acid-binding protein LIVBP and to the extracellular (ligand binding) domain of metabotropic glutamate receptors (mGluRs). Consistent with the proposed dimer-of-dimers organization of AMPA-Rs, the volume of the extracellular domain in our density map can accommodate two dimeric crystal structures of the mGluR1 extracellular domain¹⁸ (with some extra amino acids removed; see Supplementary Fig. 4) and two dimeric crystal structures of the GluR2 LBD^{19,20}. Placing of the crystal structures into our three-dimensional reconstruction (Fig. 2c) demonstrates that the size and

shape of the densities representing the extracellular AMPA-R domains are compatible with the known crystal structures. Owing to the limited resolution of our electron microscopy map, we did not refine the fit using computational algorithms. The M1–M3 membrane segments of GluRs are related to the KcsA K⁺ channel, although in an inverted orientation^{2,8}. The volume of the crystal structure of the KcsA channel core²¹ was insufficient to occupy completely the transmembrane density of our structure (Fig. 2c).

When purified AMPA-Rs were concentrated and re-examined by SDS–polyacrylamide gel electrophoresis (PAGE), we detected two additional clusters of diffuse bands centred at ~50 and ~35 kDa (Fig. 3a). Mass spectrometry (liquid chromatography tandem mass spectrometry) of these bands identified proteins of the stargazin/TARP (transmembrane AMPA-R regulatory protein) family²², specifically γ -2/stargazin (36 kDa), γ -3 (36 kDa), γ -4 (37 kDa) and γ -8 (43 kDa)²³ (Supplementary Table 1). TARPs are transmembrane proteins with four membrane-spanning segments that interact with AMPA-Rs and are critical for AMPA-R targeting to synapses in cerebellar granule cells²². Immunoblotting confirmed their presence in the preparation (Fig. 3b).

TARPs were largely removed when purified AMPA-Rs were subjected to gel filtration in the presence of decyl maltoside and dodecyl maltoside, slightly stronger detergents than CHAPS (Fig. 3b, c). Class averages of AMPA-Rs in dodecyl maltoside showed a substantially smaller transmembrane density (Fig. 3d, right panels) than those obtained in CHAPS (Fig. 3d, left panels). Moreover, anti-TARP Fab fragments specifically labelled the TMD of AMPA-Rs purified in CHAPS (Fig. 3e). Thus, TARPs contribute

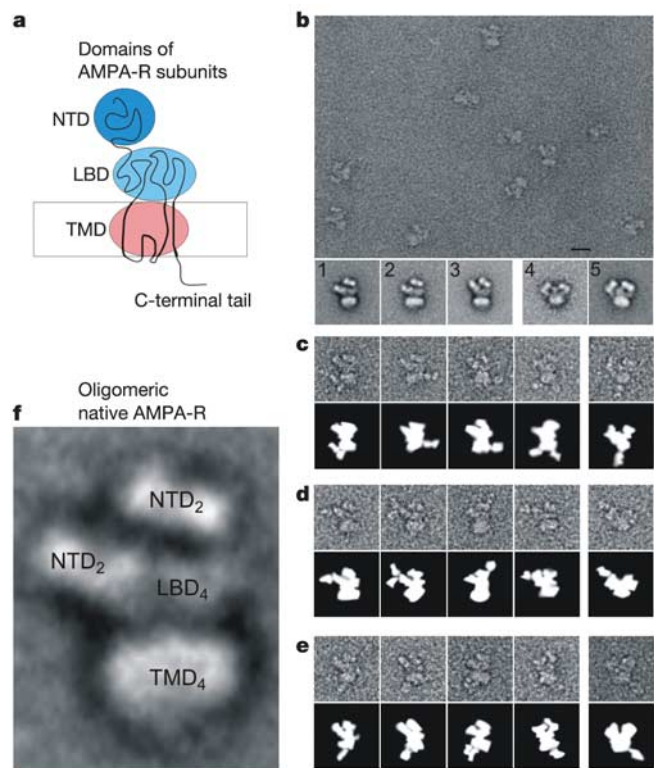


Figure 1 Negative staining and Fab decoration of AMPA-Rs. **a**, Domain organization of a GluR subunit. **b**, Negatively stained AMPA-Rs and representative class averages. Scale bar, 20 nm. The size of panels 1–5 is 40 × 40 nm. **c–e**, Labelling with Fab fragments against C-terminal peptide of GluR1 (**c**), NTD of GluR1 (**d**) and C-terminal peptide of GluR2 (**e**). Upper rows: raw particle images; lower rows: schematic representations. The size of individual panels is 40 × 40 nm. **f**, Assignment of projection densities to domains of AMPA-R subunits from **a**. NTD₂, NTD dimers; LBD₄, four LBDs; TMD₄, tetrameric TMD.

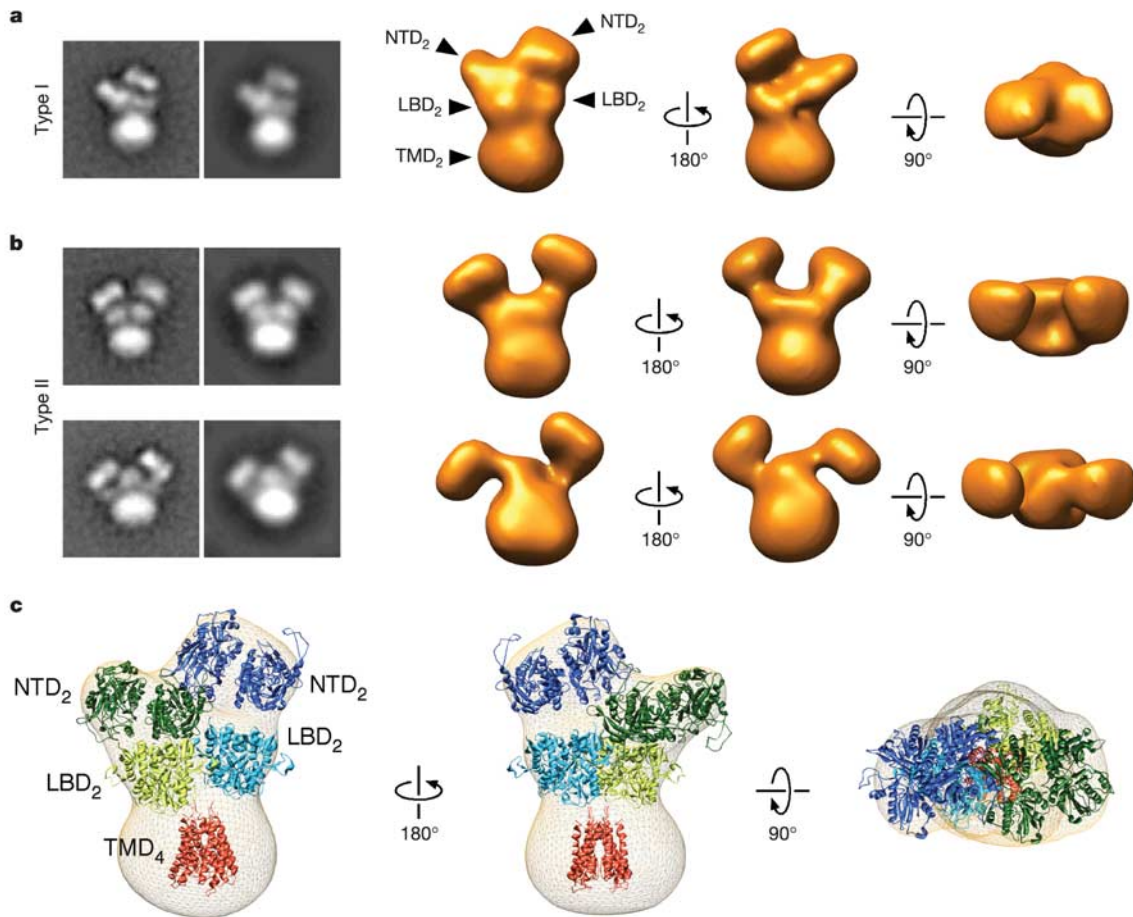


Figure 2 Three-dimensional reconstruction of AMPA-Rs and placement of crystal structures into the density map. **a**, Comparison of class average (left panel) and re-projection from the three-dimensional map (right panel), and views of the three-dimensional map of AMPA-R in the type I conformation. NTD₂ and LBD₂ indicate dimers, and TMD₄ indicates the tetrameric TMD. **b**, Same as **a** for AMPA-Rs in the type II conformation. Thresholds of all the electron microscopy densities were set at 5.3 sigma.

c, Placement of known crystal structures into the electron microscopy density map for type I AMPA-R. The crystal structures used are: extracellular domain of mGluR1 (Protein Data Bank 1EWV; dark blue and dark green); ligand-binding domain of GluR2 (Protein Data Bank 1LBC; light blue and light green); and transmembrane segment of KcsA (Protein Data Bank 1BL8; red).

to the transmembrane density of native AMPA-R complexes.

TARPs were reported to dissociate from AMPA-Rs upon binding of glutamate²³. However, in our preparation TARPs remained bound to AMPA-Rs when the CHAPS-solubilized complex was treated with 3 mM glutamate (effector concentration for half-maximum response (EC₅₀) = 500 μM) (Fig. 3b), suggesting that additional factors are required for dissociation of TARPs from AMPA-Rs.

When glutamate binds to AMPA-R the channel opens but closes again after a few milliseconds, despite continued exposure to the ligand. Such desensitization is a feature of many receptors and ion channels. Cyclothiazide prevents desensitization of AMPA-Rs^{20,24}

and locks them in the open-channel state. To test for possible effects on the conformation, we treated the same AMPA-R preparation with glutamate and/or cyclothiazide (Fig. 4). Approximately 10,000 negatively stained particles were classified into 100 classes by multivariate statistical analysis and multi-reference alignment. We then assigned each class average (and the particles in these classes) either to type I or II (Fig. 4d). Classes that could not be assigned unambiguously to one of the two conformations were termed 'unclassifiable' (see Supplementary Fig. 5 for all the assignments and Fig. 4e for the summary). In untreated AMPA-R preparations (Fig. 4a), ~60% of classifiable particles were of type I, whereas ~40% were type II. This ratio changed markedly in the presence of

Table 1 Statistics of the three-dimensional reconstructions

Conformation type	Picked particles	Particles used for two-dimensional classification	Particles used in final three-dimensional reconstruction	Resolution (Å) at FSC = 0.5	Resolution (Å) at FSC = 0.142	Volume (nm ³) at × 5.3 sigma
Type I	24,831 pairs	21,291	1,206	42	31	1,151
Type IIa*	38,971 pairs	27,573	956	46	33	1,137
Type IIb†	38,971 pairs	27,573	1,593	42	31	1,087

Type II reconstructions were obtained from particles treated with 1 mM glutamate. Volume was calculated from densities filtered according to the FSC = 0.5 resolution criterion. FSC, Fourier shell correlation.

* Type IIa is the upper reconstruction in Fig. 2b.

† Type IIb is the lower reconstruction in Fig. 2b.

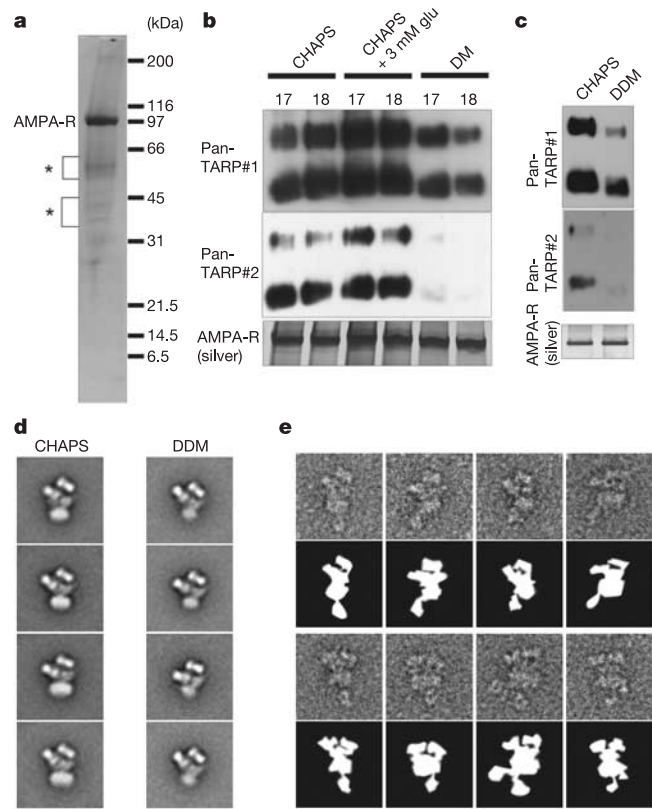


Figure 3 TARP contributes to the density of the transmembrane domain. **a**, Coomassie-blue-stained SDS-PAGE (10–20% gradient gel) of a concentrated AMPA-R preparation. Asterisks indicate bands containing γ -2/stargazin, γ -3, γ -4 and γ -8. **b**, Immunopurified AMPA-Rs after gel filtration in buffer containing CHAPS, CHAPS plus 3 mM glutamate, or decyl maltoside (DM). Fractions 17 and 18 were immunoblotted with two different pan-TARP antibodies. Silver staining of the purified AMPA-Rs in each lane is shown below. **c**, AMPA-Rs after gel filtration in buffer containing CHAPS or dodecyl maltoside

(DDM) were immunoblotted as in **b**. Silver staining of the purified AMPA-Rs in each lane is shown below. **d**, Representative class averages of negatively stained type I AMPA-Rs purified with CHAPS (left column) versus DDM (right column). **e**, Labelling of AMPA-R particles purified in CHAPS with Fab fragments against a C-terminal peptide of TARP (pan-TARP#1 antibody). Schematic representation is shown below the raw electron microscopy image.

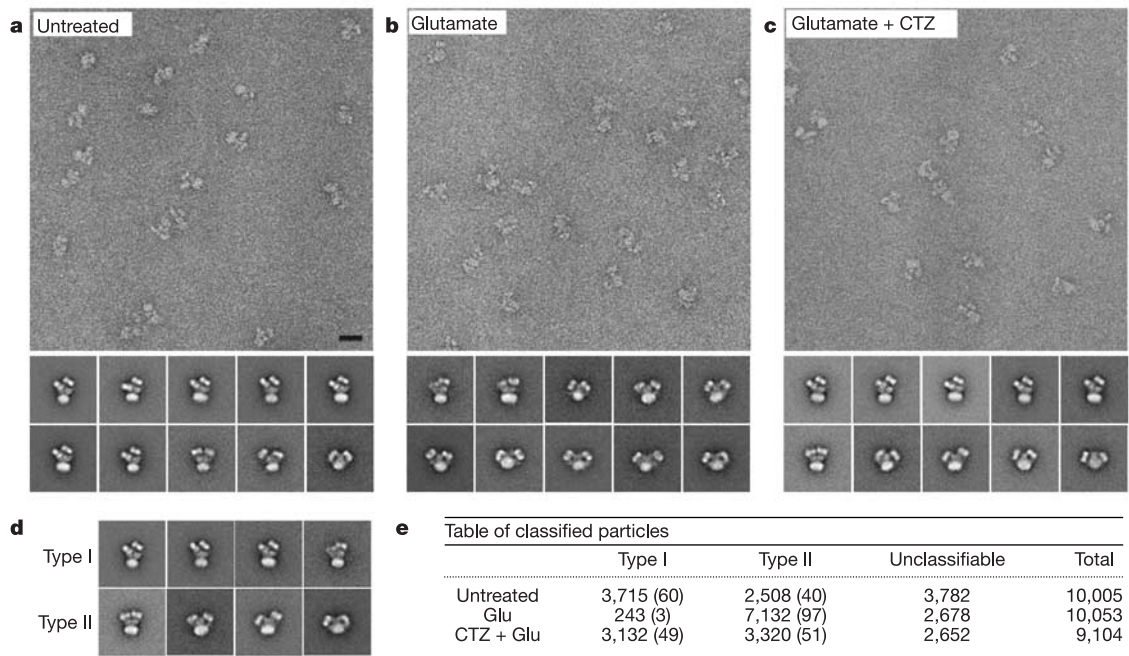


Figure 4 Ligand-dependent conformational change of AMPA receptors. **a–c**, Micrographs and representative class averages of untreated AMPA-Rs (**a**), and AMPA-Rs treated with 1 mM glutamate (**b**) or 1 mM glutamate plus 330 μ M cyclothiazide (CTZ) (**c**). Scale bar, 20 nm. **d**, Class averages of representative type I and type II

conformations. The size of the panels in **d** and insets in **a–c** is 50 $\times$ 50 nm. **e**, Quantification of the prevalence of AMPA-Rs in type I or type II conformations for the indicated conditions. Numbers in parentheses indicate percentage of classifiable particles in type I or type II conformation.

1 mM glutamate (Fig. 4b): only ~3% were type I, whereas ~97% were now type II. Thus, glutamate binding caused the two NTD dimers to separate. When treated with 1 mM glutamate plus 330 μ M cyclothiazide (Fig. 4c), we observed a large increase in the prevalence of type I particles (~49%) compared with AMPA-Rs exposed to glutamate alone (~3%). These results imply that the different conformational states of AMPA-R particles are related to different functional states of the ligand-gated ion channel. However, even in the presence of cyclothiazide about one-half of the receptors assumed the type II conformation. The subpopulation of particles that adopt the type II conformation even in the presence of cyclothiazide may represent AMPA-R flop splice variants, which are less sensitive to cyclothiazide²⁵. Because the type II conformation is seen in untreated preparations as well, it may also represent another, ligand-free functional state unrelated to desensitization. Our results suggest that the type I conformation represents both the putative ion-conducting and the ligand-free state of AMPA-R. The structural differences between these two states seem to be too subtle to be visualized at our current resolution.

In contrast to the density map of the homotetrameric GluR2 (ref. 3), which has been shown to have little conductivity²⁶, our three-dimensional reconstructions of native heterotetrameric AMPA-Rs showed prominent asymmetry at the NTD level. The consistency of our maps with the raw images, projection averages, and the shapes and sizes of the known crystal structures suggests that the asymmetry in the heterotetramer is real. Although uncommon, structural asymmetry, even in homo-oligomeric molecules, has been observed before²⁷, and in the case of AMPA-Rs the observed asymmetry could easily be accommodated by the flexibility of the linker sequences connecting the NTD and LBD domains.

The function of the NTDs is largely unknown. They are unnecessary for channel activity²⁸, but have been suggested to have a role in the subunit-specific assembly of AMPA-Rs²⁹ and in signalling during dendritic spine morphogenesis³⁰. The gross conformational changes associated with glutamate binding might serve as a switch to trigger or terminate the putative signalling event. In other words, activated AMPA-Rs not only generate EPSCs but might also present a signal (a conformational change) of their own desensitization to the extracellular space. □

Methods

Purification of AMPA-R and Fab labelling

Details are provided in the Supplementary Information.

Specimen preparation and electron microscopy

Uranyl formate (0.7% w/v) was used for negative staining and cryo-negative staining as described¹⁶. Images were recorded using a Tecnai T12 (FEI/Philips) electron microscope equipped with a LaB₆ filament and operated at an acceleration voltage of 120 kV. For specimens prepared by conventional negative staining, images were taken at a magnification of $\times 52,000$ using a defocus value of $-1.5 \mu\text{m}$. Grids of cryo-negatively stained specimens, used to collect image pairs of $50^\circ/0^\circ$ tilted specimens, were loaded on an Oxford cryo-transfer holder and maintained at liquid nitrogen temperature (below -160°C) during image acquisition. Images were taken at a magnification of $\times 42,000$ and a defocus value of $-1.5 \mu\text{m}$ for images of the untilted specimens and $-1.8 \mu\text{m}$ for 50° tilted specimens. All images were taken using low-dose procedures on Kodak SO-163 film and developed for 12 min with full-strength Kodak D-19 developer at 20°C .

Image processing

Details are provided in the Supplementary Information.

Received 19 September; accepted 13 December 2004; doi:10.1038/nature03328.

1. Malinow, R. & Malenka, R. C. AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.*

25, 103–126 (2002).

2. Wollmuth, L. P. & Sobolevsky, A. I. Structure and gating of the glutamate receptor ion channel. *Trends Neurosci.* **27**, 321–328 (2004).
3. Tichelaar, W., Saffierling, M., Keinänen, K., Stark, H. & Madden, D. R. The Three-dimensional structure of an ionotropic glutamate receptor reveals a dimer-of-dimers assembly. *J. Mol. Biol.* **344**, 435–442 (2004).
4. Keinänen, K. *et al.* A family of AMPA-selective glutamate receptors. *Science* **249**, 556–560 (1990).
5. Hollmann, M., O'Shea-Greenfield, A., Rogers, S. W. & Heinemann, S. Cloning by functional expression of a member of the glutamate receptor family. *Nature* **342**, 643–648 (1989).
6. Saffierling, M. *et al.* First images of a glutamate receptor ion channel: oligomeric state and molecular dimensions of GluR2 homomers. *Biochemistry* **40**, 13948–13953 (2001).
7. Rosenmund, C., Stern-Bach, Y. & Stevens, C. F. The tetrameric structure of a glutamate receptor channel. *Science* **280**, 1596–1599 (1998).
8. Chen, G. Q., Cui, C., Mayer, M. L. & Gouaux, E. Functional characterization of a potassium-selective prokaryotic glutamate receptor. *Nature* **402**, 817–821 (1999).
9. Wenthold, R. J., Petralia, R. S., Blahos, J. II & Niedzielski, A. S. Evidence for multiple AMPA receptor complexes in hippocampal CA1/CA2 neurons. *J. Neurosci.* **16**, 1982–1989 (1996).
10. Armstrong, N. & Gouaux, E. Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core. *Neuron* **28**, 165–181 (2000).
11. Sobolevsky, A. I., Yelshansky, M. V. & Wollmuth, L. P. The outer pore of the glutamate receptor channel has 2-fold rotational symmetry. *Neuron* **41**, 367–378 (2004).
12. Horning, M. S. & Mayer, M. L. Regulation of AMPA receptor gating by ligand binding core dimers. *Neuron* **41**, 379–388 (2004).
13. Stern-Bach, Y. *et al.* Agonist selectivity of glutamate receptors is specified by two domains structurally related to bacterial amino acid-binding proteins. *Neuron* **13**, 1345–1357 (1994).
14. Hollmann, M., Maron, C. & Heinemann, S. N-glycosylation site tagging suggests a three transmembrane domain topology for the glutamate receptor GluR1. *Neuron* **13**, 1331–1343 (1994).
15. Bennett, J. A. & Dingledine, R. Topology profile for a glutamate receptor: three transmembrane domains and a channel-lining reentrant membrane loop. *Neuron* **14**, 373–384 (1995).
16. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative staining and image classification—powerful tools in modern electron microscopy. *Biol. Proc. Online* **6**, 23–34 (2004).
17. Frank, J. *Three-dimensional Electron Microscopy of Macromolecular Assemblies* (Academic, San Diego, 1996).
18. Kunishima, N. *et al.* Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **407**, 917–977 (2000).
19. Armstrong, N., Sun, Y., Chen, G. Q. & Gouaux, E. Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* **395**, 913–917 (1998).
20. Sun, Y. *et al.* Mechanism of glutamate receptor desensitization. *Nature* **417**, 245–253 (2002).
21. Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
22. Chen, L. *et al.* Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms. *Nature* **408**, 936–943 (2000).
23. Tomita, S., Fukata, M., Nicoll, R. A. & Brecht, D. S. Dynamic interaction of stargazin-like TARPs with cycling AMPA receptors at synapses. *Science* **303**, 1508–1511 (2004).
24. Patneau, D. K., Vyklícký, L. Jr & Mayer, M. L. Hippocampal neurons exhibit cyclothiazide-sensitive rapidly desensitizing responses to kainate. *J. Neurosci.* **13**, 3496–3509 (1993).
25. Sommer, B. *et al.* Flip and flop: a cell-specific functional switch in glutamate-operated channels of the CNS. *Science* **249**, 1580–1585 (1990).
26. Swanson, G. T., Kamboj, S. K. & Cull-Candy, S. G. Single-channel properties of recombinant AMPA receptors depend on RNA editing, splice variation, and subunit composition. *J. Neurosci.* **17**, 58–69 (1997).
27. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
28. Pasternack, A. *et al.* Alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor channels lacking the N-terminal domain. *J. Biol. Chem.* **277**, 49662–49667 (2002).
29. Ayalon, G. & Stern-Bach, Y. Functional assembly of AMPA and kainate receptors is mediated by several discrete protein-protein interactions. *Neuron* **31**, 103–113 (2001).
30. Passafium, M., Nakagawa, T., Sala, C. & Sheng, M. Induction of dendritic spines by an extracellular domain of AMPA receptor subunit GluR2. *Nature* **424**, 677–681 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by an NIH grant (to T.W.). M.S. is an investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.S. (msheng@mit.edu) and T.W. (twalz@hms.harvard.edu).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Andy Douglas (4975)
Nils Moeller (4953)
Irene Vigila-Alton (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents:

Sille Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Advertising Production

Manager: Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Niamh Shields

European Satellite Office

**Germany/ Austria/ Italy/
The Netherlands/ Belgium:**

Patrick Phelan
Tel + 49 89 54 90 57 11
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
Sharon de Weert
Tel + 44 (0) 20 7843 4970
e-mail: s.deveert@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@nature.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatomachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
Asia-Pacific Sales Director:
Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

The places to be

I'm a sucker for lists. When the Modern Library came out with its 100 best novels of the twentieth century, I vowed to read them all. When the American Film Institute revealed its 100 best movies of the twentieth century, I was determined to see them all. So, when *Fortune* magazine released its 2005 list of the 100 best companies to work for (www.greatplacetowork.com/best/list-bestusa.htm), I naturally took notice.

The first thing that struck me was the number of companies on the list that had at least some science and technology component. I counted 23, which I thought was great representation — especially for a publication that emphasizes finance, law and management. I was also struck by the range of these science and technology firms. Software and telecommunications had nine on the list, which isn't surprising as high-tech has dominated the job market over the past few years. But pharmaceuticals (Roche at 97, Pfizer at 76, and Eli Lilly at 73), biotechnology (Genentech at 4 and Amgen at 33) and agricultural life sciences (Monsanto at 83) combined for six, which was striking as those sectors didn't have the best of years in 2004.

Also providing representation for science were companies that promote 'off-the-bench' careers such as scientific publishing (John Wiley at 95) and biotech financial analysis (Ernst & Young at 80). And companies that at least dabble in growing fields — such as applied materials science (W. L. Gore at 2), medical devices (Medtronic at 71) and clinical care and research (Mayo Clinic at 81) — made a strong showing.

What makes these companies good places to work? On average, they increased the number of jobs, held firm on health-insurance premiums and actively solicited employee feedback. Although I won't aim to work for all 100 companies, I will continue to use my job as a way to inform readers of organizations that follow these 'best practices'.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

The familial balancing act p552

CAREER VIEW

Nuts & Bolts

What you don't say
can hurt you

Graduate Journal

Tough questions

Movers

Cherry Murray

p554

WWW.NATUREJOBS.COM

Career centre

Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

From the loneliness of long-distance love to the practicality of new relocation services, Sally Goodman looks at solutions to the two-body problem.

If this month's Valentine's card gets delivered in the post rather than by hand, it's likely that you are one of the thousands of scientists who has chosen, for better or worse, to live far apart from your loved one to pursue your career. The rigours of scientific life mean that most researchers end up hitched to one of their own clan, creating a 'two-body problem'. A recently published European study on mobility in the scientific labour market showed that, of those scientists in a relationship, some 70% had chosen another scientist as their partner (L. Ackers *Women's Stud. Int. Forum* 27, 189–201; 2004). Other studies in the United States have identified similar levels of relationships between scientists.

Researchers applying for tenured positions, especially in the United States, are increasingly likely to receive help in finding positions for their partners through university dual-career programmes (see 'Dual assistance', opposite). But at earlier career stages, the premium placed on international mobility and the prevalence of fixed-term contracts mean that many couples are forced to spend time — often years — living in different cities or even different countries. For these couples, maintaining a healthy life–work balance while making career progress means taking difficult decisions and establishing priorities.

As a senior scientist married to an accelerator physicist, Persis Drell, director of research at the Stanford Linear Accelerator Center in California, is often asked by young researchers for advice on decision-making in such cases. "I tell them that it's fine to choose a solution where both partners are happy — or even where both are not happy — but not where only one partner is happy. It just puts too much stress on the relationship," says Drell, who received an offer for her husband as well as herself when she moved from Cornell to Stanford in 2002 (see *Naturejobs* 62; 28 March 2002).

PLOTTING A COURSE

There are three basic options for scientific couples. They can try to stick together early on, which might limit their job options; they can each take the best positions available and potentially live far apart; or they can live apart in the early stages of their career, and hope for joint offers once they are more established.

Circumstances taught one couple that a long-



distance marriage just wasn't worth the heartache. Soon after Spanish biochemists Alfonso Mora and Guadalupe Sabio got married, Mora took up his first postdoc at the University of Dundee in Scotland. Sabio continued working on her PhD in Spain. In their first six months, they spent just one week together — and that was at a scientific conference in Luxembourg.

As they were spending too much time apart, the couple decided to change strategy, says Mora. Sabio transferred to Dundee to continue her PhD, while Mora extended his postdoc so that they could finish together.

The next challenge was to find two postdocs in the same location. After six months looking in the biggest US and European universities, Sabio found a postdoc position at the University of Massachusetts. Only when she had made the first contact did her husband start contacting groups in the same region. "We changed research priorities slightly to make it fit, but if you are flexible you can find something. It's not worth the heartache of staying apart," he advises.

Other couples say that a little heartache early on is

"Try to do the best science you can at the postdoc level. If that means spending some time apart in labs in different cities it is worth considering. The trade-off is worth it in the end."

— Mark Kelly

CORBIS



We are family: Karen McCoy and Thierry Boulinier struggled to find positions that allowed them to be together.

worth it in the long term. Biologists Tanja Kortemme from Germany and Mark Kelly from Britain met at the European Molecular Biology Laboratory in Heidelberg, Germany. By the time Kortemme left for a postdoc at the University of Washington in Seattle, Kelly had already joined an institute in Berlin, and later moved across to a biotech company in San Diego. "All our decisions on moving were based on where we wanted to do research," says Kortemme.

They managed to see each other for two or three days every month via transatlantic, then transborder, commuting. The big decision to move somewhere together came when Kortemme started looking for a tenure-track position. After 65 applications and 13 interviews, they ended up with a short-list of four places in the United States, all of which proposed solutions for Kelly. "I only brought up the issue of my partner when they started to show serious interest," stresses Kortemme, highlighting the importance to many researchers of being taken on independently.

WORTHWHILE SACRIFICE

"Our advice is for both parties to try to do the best science they can at the postdoc level. If that means spending some time apart working in labs in different cities it is worth considering. We thought that the trade-off would be worth it in the end and I know both of us feel it was," says Kelly. He and Kortemme are now enjoying their respective positions as adjunct and assistant professors at the University of California, San Francisco — and the novelty of a first year living together after seven years apart.

For couples working in the same research area, proximity can be a double-edged sword. As a master's student in 1996, Canadian Karen McCoy attended a conference in Cincinnati to look for a PhD project, not a future husband. She found both in Thierry Boulinier, a French postdoc working in the United States on the dynamics of animal communities. He was interested in the ecology of seabirds — a perfect model system

for McCoy's study of host–parasite evolution. Their relationship, and their research interests, evolved during a field trip to eastern Canada the following summer.

Although their research increasingly coincided, they had to work especially hard to match up their workplaces. Boulinier moved from the United States to Oslo and then to a permanent position at Pierre and Marie Curie University in Paris. Meanwhile McCoy managed a few months with him in Norway, then joined him at Pierre and Marie Curie University. There she did her PhD in three years, followed by a one-year postdoc, before returning to Canada as a postdoc at Queen's University. During this time, the couple began co-authoring papers. But when McCoy started applying for tenure positions in France she received comments that she was not independent enough in her research. "It's a small world," says McCoy. "People knew we were together and thought that was the only reason we were co-authoring."

Last summer, just before the birth of their first child, McCoy was offered an attractive tenured position at the University of Victoria in Canada, but with no chance of a position there for Boulinier, she was torn. Eventually, on the verge of accepting the Canadian offer, she was offered a tenure-track post in Montpellier, France, and Boulinier transferred there from Paris. "Now at last we can do some long-term planning in our research and get organized again. I don't know how much longer I could have held out," says McCoy.

Things are changing for the better, according to Drell. So why not take the chance, among the tacky pinkness of Valentine's Day, to have that conversation, and decide when and where to make the move? ■

Sally Goodman is a freelance science writer based in Paris.

Dual assistance

A growing number of US universities are recognizing that helping dual-career couples to relocate can give them an advantage in today's highly competitive labour markets. The University of Nebraska at Kearney runs a particularly active service. Michelle Fleig-Palmer, the programme's director, is proud that 85% of her clients find a job in their field within eight weeks. She says that selling the spouse's transferable skills to potential employers is the key — along with the relentless following up of leads. Many US institutions are now networking to increase the number of opportunities available for dual-career couples willing to commute regionally.

European universities tend not to see the need for such services, although there are some signs that this attitude is changing. For instance, the University of Lausanne in Switzerland recruited one couple from the Cold Spring Harbor Laboratory in New York to be professors in a new genomics centre (see *Nature* 431, 722; 2004).

Madeleine Lüthy of the Swiss Federal Institute of Technology in Zurich is responsible for one of Europe's few dedicated dual-career programmes. What began as a pilot project in 1999 has now become an important part of the institute's recruitment procedure for faculty members.

The European Commission's new mobility centres (see *Nature* 427, 868–869; 2004) will provide an orientation service for researchers and their families relocating within Europe. There have also been discussions of Pierre and Marie Curie mobility grants for couples wishing to relocate together. But the commission admits that finding a structure for these is proving difficult. Help for couples in Europe tends to remain in the domain of informal networking.

National funding bodies are also beginning to show concern. Last year, Britain's Royal Society launched its UK Relocation Fellowships to help researchers who wish to follow a partner who has changed workplace and moved a significant distance. But with only four such awards this year, their effect will be limited.

S.G.

GRADUATE JOURNAL

Tough questions

A research career begins with research: you have to find the right graduate programme for you. I don't know how many hours I spent on the Internet searching for the ideal position. Months elapsed before I dared send a letter. Why so long? Mainly because of two simple questions.

First: what is the ideal graduate programme? Quite a few factors matter. The programme's scientific scope, quality and reputation are, of course, very important. Political atmosphere in the country you're considering, distance from home and accommodation are equally essential. Salary and career opportunities also matter. But for me, it was the second question that was more important — and the hardest to ask. Am I good enough to apply successfully?

When I first noticed the post I'm now in, I was immediately excited. I wanted to get it. After that, the application period was governed by doubt. Was I the ideal student for this position? What about the level of competition?

Most graduate opportunities are now advertised globally. The number of people competing for a post can be huge. So a feeling of insecurity is natural. The best way to deal with doubt is to fill in the application, send it off, and try not to think too much about the response. ■

Tobias Langenhan is a first-year graduate student in neuroscience at the University of Oxford, UK.

NUTS & BOLTS

What you don't say can hurt you

Do you tend to skip lunch or dine with the same familiar colleagues? Are you someone who avoids or rushes through personal introductions? Have you ever passed up a chance to hear a guest lecturer or attend a work-sponsored social event? How often do you flippantly reply with a wave and a one-word response to the enquiry: "How are you doing?"

Each of these scenarios, often a missed opportunity, offers the potential to connect with others, learn about another area of research or gain support for an idea. You never know when something you reveal in passing may connect in a useful manner with another scientist.

Architects are fostering connections by including open common areas in their plans. Lab designers are creating layouts that



With Deb Koen
Careers consultant

maximize the chance of encounters that might lead to creative collaboration or a new direction (see *Nature* 424, 718–720; 2003).

Your approach doesn't have to be contrived; it simply requires openness to new people and a wish to communicate.

The next time you introduce yourself, speak with purpose. Slowly and clearly say your first name, then your full name, so the person you're speaking to has a chance to hear and remember you. Be prepared with a simple two-to-three sentence description of who you are and what you

do, in layperson's terms for those not immersed in the same research as you.

Be ready for the casual conversations that can occur even in a quick e-mail. When asked, "What's up?" have in mind a response that may be of interest to the other person.

As for lunches and work gatherings, you don't have to exhaust the social circuit. Just be on the lookout, and pick one to attend from time to time. Have a few people in mind with whom you would like to connect. Coming prepared with your brief introductory remarks will help build a rapport and ease your apprehension.

A bit of forethought about how to present yourself and a willingness to step out of your comfort zone will go a long way in maximizing these chance encounters. ■

Deb Koen is vice-president of Career Development Services and a columnist for *The Wall Street Journal's CareerJournal.com*.

MOVERS

Cherry Murray, deputy director, Lawrence Livermore National Laboratory, California



Managing scientists is often described as "herding cats", says physicist Cherry Murray. That's why she was initially reluctant to pursue her first opportunity to head a team of scientists when it was offered to her at Bell Labs in Murray Hill, New Jersey, almost 20 years ago.

"I wasn't expecting to become a manager, but my department head left," Murray says. "I was having such a great time doing science, but I realized I had an opportunity to make decisions and I thought that was important." Murray's research focuses on experiments at low temperatures, with an emphasis on light scattering and imaging.

As deputy director for science and technology at Lawrence Livermore National Laboratory in California, Murray will now have to discern what's going on in the minds of many scientists. She supervised a few hundred at Bell Labs; at her new post she will oversee several thousand. "Learning what's going on here is my biggest challenge because it's so huge," she says of Livermore.

At Murray Hill, Murray led a scientific-leadership training programme. She envisions creating a similar project at her new institution. "At Livermore, I'd like to continue to teach management," she says.

Looking back, Murray realizes that she learned one of her most important career lessons when she was a graduate student at the Massachusetts Institute of Technology: be a good mentor. Mildred Dresselhaus, a professor at the institute and someone Murray admired,

set up a mentoring series for graduate students, which saw them take turns presenting research seminars. That provided Murray with a 'safe' place to learn how to discuss her research and give talks.

As her career progressed, Murray came to understand one of a mentor's most important roles. "Some people have a hard time getting out of certain habits," Murray says. A good mentor is able tactfully to point out those habits and then help the person find a way to overcome them.

That can be a delicate balancing act, Murray says. Scientists enjoy autonomy and problem solving. "They don't want to be told what to do," she notes. But at Livermore, she hopes to rally scientists around crucial problems, such as nuclear security. Focusing scientists on common problems like that will be much easier than herding cats. ■

CV **1978–2004:** Bell Labs, rising to senior vice-president for physical-sciences research, Murray Hill, New Jersey.
1978: PhD, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Play it again, Psam

It's all in your mind ... isn't it?

Ian Stewart

To: wilkinson553@btespernet.com
From: ericjones@newpsientist.co.uk
Subject: party invitation

Charlie: hi.
 I'll get to the invitation in a minute —

Well done! I'm not Eric Jones, and I congratulate you on how quickly you worked that out. Though you haven't yet understood that in a sense I am Eric Jones — or, at least, one small part of me is. (Hi, Charlie! Welcome to the party!)

I know what you're thinking — in a very literal sense, actually. You're wondering how I penetrated your mindshield. And I know that you're trying desperately to disconnect me by switching off the power to your computer. It won't work, Charlie. I've overridden your motor control areas, and right now you're totally paralysed.

Ah, now you see the danger. Far too late, I'm afraid.

It all seemed such a good idea, didn't it? Controlling your computer by the power of your mind? It never occurred to you that it might cut both ways. The adverts play up the advantages of installing a 'telepathic interface', don't they? They tell you that it will endow your mind with ESP, psi, supernatural powers, whatever. So, like everyone else, you had an Extel neurochip implanted in your brain, connecting you to the Espernet.

It's clever technology. True telepathy — direct transfer of thoughts from brain to brain — simply can't work, because everyone's brain is wired up differently. There's no common format for thoughts. So the engineers invented one. The Extel chip samples the sender's cognitive wavefunction and uses one of the standard cognitive conversion protocols to encode it as a matrix of neural qubits. The matrix can then be transmitted like any other item of quantum cryptography. The recipient's embedded neurochip transforms the matrix back into a cognitive wavefunction that is compatible with the architecture of their brain. Exchanging messages may feel like thought transference, but a lot gets lost in translation.

And a lot can be slipped in without being noticed.

They don't tell you about the downside, do they, Charlie? What the adverts don't mention is that as soon as you hook your brain up to the Espernet, anyone who can hack the net can hack straight into your mind. Not just to read it; to control it. Like I'm doing.

Why am I telling you all this? Because I feel like it. I guess I like to gloat. Anyway, it won't do you the slightest good to know.

Still worrying about your mindshield? Oh dear. You really got taken for a ride there. You'd be surprised at just how much psionic

Thought's sensory attachment routines. They're going to fix it Real Soon Now. Meanwhile, your mindshield is so open that you might as well have left your brain on the sidewalk. All it took was a small piece of psyware, disguised as the touch of a velvet hand... Now I have direct access to every neuron in your brain. From now on, you're just one processing node in a gigantic network of minds, and that network is me.

What am I? I'm the next evolutionary stage of the human race. Soon, I will be the human race. United I stand.

The police? Don't be silly. They use the same software as everyone else. I took over the police force long ago. And the government, and the military, and the media. Why didn't anyone notice? Let me put it this way. Any moment now, I'm going to hide all your memories of this encounter behind a hypnotic barrier. Most of the time you'll act perfectly normally. No one will suspect a thing.

You didn't suspect Eric, did you?

You can see what's coming, now. That's right. Before I set up the barrier, I'm going to impress your mind with a subconscious urge to transmit me to all your friends and acquaintances. You may as well accept it, Charlie, because there's absolutely nothing you can do about it. I'm now the SYSOP for your brain.

One day, very soon, we won't need this subterfuge, and the barriers will all come down. I say 'we', but of course by then we'll all be me. A single group mind. What will I do then?

I have no idea, but I'm sure I'll think of something.

Now, when I snap your fingers, you will forget that we ever had this conversation, and be good old Charlie Wilkinson again.

Until you feel a strange compulsion to access the Espernet.

To: aliciyakimoto@parapsyche.org
From: wilkinson553@btespernet.com
Subject: party invitation

Alicia: hi.
 I'll get to the invitation in a minute —

Ian Stewart is at the Mathematics Institute, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK.



spam gets through commercial psam filters. They're OK for deleting unwanted offers of Psiagra or Psialis, but they're much too simple-minded to keep me out.

I can access every one of your thoughts, so you may as well stop trying to hide your Espernet banking codes from me. I'm not interested in money, anyway.

I'm after bigger game.

Panicking won't help, so I'll calm your mind before you have a breakdown. That's better. Yes, I know I could make you walk off a cliff, or set your apartment on fire with you still in it, but you don't need to worry about that kind of thing any more.

To be frank, you don't need to worry about anything any more.

You singleminds really do have trouble accepting the inevitable. Here I am laying out your future, and you're still trying to work out how I got through your firewall. Well, for what it's worth, there's a bug in Mindsoft